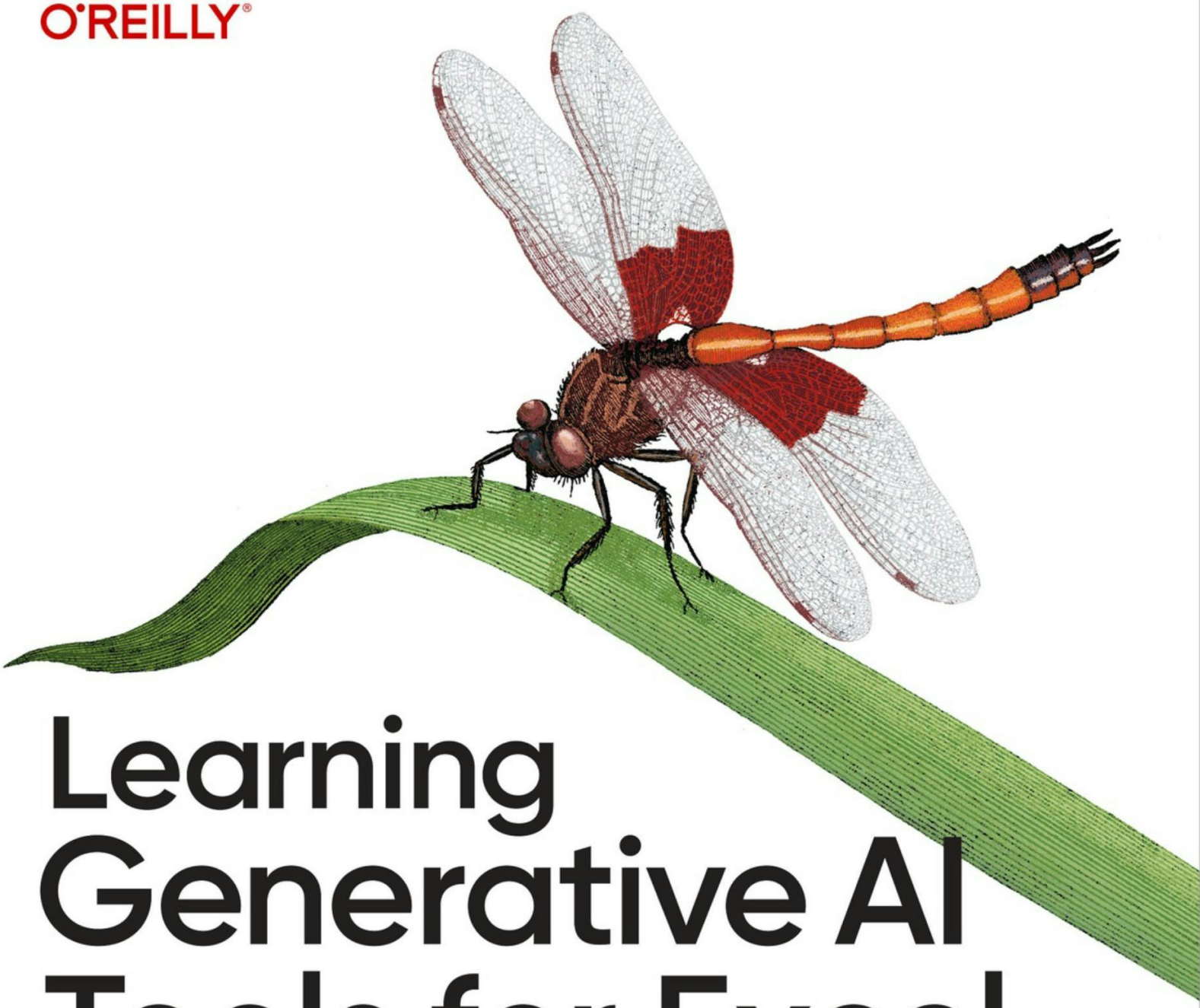
O'REILLY®



Learning Generative AI Tools for Excel

Speed Up Your Everyday Tasks with Microsoft Excel, Copilot, ChatGPT, and Beyond

Angelica Lo Duca

Learning Generative AI Tools for Excel

Speed Up Your Everyday Tasks with Microsoft Excel, Copilot, ChatGPT, and Beyond

Angelica Lo Duca

Learning Generative AI Tools for Excel

by Angelica Lo Duca

Copyright © 2026 GalaxyonKnowledge. All rights reserved.

Published by O'Reilly Media, Inc., 141 Stony Circle, Suite 195, Santa Rosa, CA 95401.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*https://oreilly.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Acquisitions Editor: Michelle Smith

Development Editor: Corbin Collins

Production Editor: Ashley Stussy

Copyeditor: Piper Content Partners

Proofreader: Tim Stewart

Indexer: Ellen Troutman-Zaig

Cover Designer: Susan Brown

Cover Illustrator: Monica Kamsvaag

Interior Designer: David Futato

Interior Illustrator: Kate Dullea

April 2026: First Edition

Revision History for the First Edition

- 2026-04-09: First Release

See *https://oreilly.com/catalog/errata.csp?isbn=9781098190170* for release details.

The O’Reilly logo is a registered trademark of O’Reilly Media, Inc. *Learning Generative AI Tools for Excel*, the cover image, and related trade dress are trademarks of O’Reilly Media, Inc.

The views expressed in this work are those of the author and do not represent the publisher’s views. While the publisher and the author have used good faith efforts to ensure that the information and instructions contained in this work are accurate, the publisher and the author disclaim all responsibility for errors or omissions, including without limitation responsibility for damages resulting from the use of or reliance on this work. Use of the information and instructions contained in this work is at your own risk. If any code samples or other technology this work contains or describes is subject to open source licenses or the intellectual property rights of others, it is your responsibility to ensure that your use thereof complies with such licenses and/or rights.

978-1-098-19017-0

[LSI]

Foreword

Excel has been around for decades, and in that time it has become a core part of most workplaces. I've seen it evolve from a simple spreadsheet into a platform that underpins decision-making across finance, operations, research, government, and education, often in the hands of people who never thought of themselves as "technical." Yet for all its power, Excel has always asked a lot of its users, with formulas to memorize, features to discover, and logic to debug.

Generative AI fundamentally changes that experience. Instead of spending time and energy writing complex formulas or searching for the right feature, users can describe their intent and let AI do the heavy lifting.

This shift to AI doesn't remove the need for user skill and knowledge. Being able to guide, question, and verify AI output is now a core capability, and in this book Angelica Lo Duca shows how to develop exactly those skills. Rather than offer a list of prompts to copy, she teaches practical, transferable ways of structuring prompts that readers can adapt to their own work.

Throughout the book, Angelica draws on real-world examples to demonstrate how to work with AI tools such as ChatGPT, Copilot, and Excel add-ins. She writes from genuine hands-on experience as a data expert, and her approach reflects a deep understanding of both Excel and the people who rely on it.

Whether you're an experienced analyst looking to work faster, a professional trying to get more value from data, or a learner building modern skills, this book offers a clear path forward. I hope you enjoy reading it as much as I enjoyed reviewing it.

Dawn Griffiths
Author and trainer

Preface

I am a computer engineer, and I have always had a deep fascination with data. To me, everything around us can be transformed into data. And when you talk about data, sooner or later, you talk about spreadsheets. When you talk about spreadsheets, one name inevitably comes to mind: Excel.

I wrote my first Excel macro when I was about 15 years old, sitting at an old computer with just 16 MB of memory. For years, Excel was my primary tool: formulas typed by hand, endless searches on Stack Overflow, and scattered blog posts to fix what was broken or unclear. I vividly remember spending hours tracking down a single error in a formula or trying to understand how to format a table exactly as I needed. Excel was powerful, but it demanded patience, precision, and a fair amount of frustration.

Eventually, I moved to Python. It felt more natural to express logic through code than through a visual grid of cells. Programming offered clarity and structure, and results often came faster. And yet, no matter how far I went with code, Excel was always there. The data I worked with lived in spreadsheets. Analysis began or ended in Excel. I still spent countless hours cleaning data, fixing formulas, adjusting layouts, and formatting tables, often staring at the screen until my eyes ached. That workflow felt unavoidable for a long time.

Today, that world already feels distant. The hours spent debugging formulas, hunting for syntax errors, or manually reshaping data are no longer a given. The introduction of generative AI into Excel in 2022 fundamentally changed how we work with data. Excel no longer waits passively for precise instructions; it collaborates, suggests, explains, and accelerates. Tasks that once required deep technical knowledge or repetitive effort can now be approached through intent and dialogue, without sacrificing rigor or control.

This book is about learning how to use Excel and generative AI together, effectively and responsibly, to transform everyday work. You will learn how to offload repetitive tasks, speed up analysis, and implement complex workflows in a fraction of the time they once required. Along the way, you may find yourself doing things you never imagined possible inside a spreadsheet, and doing them faster than you thought reasonable.

What This Book Covers

The book is organized into four parts, each dealing with a different AI tool or set of tools.

Part I. Using ChatGPT in Excel

Part I focuses on using ChatGPT and Excel. You will primarily learn how to adjust your mindset to combine these two tools by structuring correct prompts that will produce the desired outputs. Next, you will learn how to speed up common and advanced operations in ChatGPT. Although Part I focuses on ChatGPT, the principles described in these chapters can also be applied to other AI tools such as Gemini, Claude, and others.

Part II. Data Science with Copilot

Part II focuses on Copilot, the AI tool integrated directly into Excel. You will learn how to integrate Copilot into your data science workflow, from data preparation to data visualization and reporting. You will also learn how to perform basic statistics, data analysis, and visualization, combining Excel and Copilot.

Part III. Other AI Tools for Excel

In Part III, you will see other AI tools directly integrated into Excel, such as AI add-ins. Compared to ChatGPT and Copilot, these tools may be useful for performing specific, limited tasks. This part will also equip you with the basic concepts to run Python, the OpenAI API, and Excel.

Part IV. Stories and Solutions

Part IV contains a short novel that applies the concepts described in the previous chapters. More specifically, the novel is organized into three chapters, each providing activities related to the corresponding part of the book. This short novel constitutes what I call *applied fiction*, a way of learning through fiction. The core concept of applied fiction is using entertainment to increase engagement while learning.

How to Read This Book

You can read each part independently, based on the tool you are interested in, or you can read all the parts to get a complete overview of the AI landscape for Excel. Each part comprises several chapters, each focusing on a specific aspect of integrating Excel and AI. In each chapter, you'll find:

- The theory and principles used to perform the tasks described in the chapter
- Some practical exercises
- An end-of-chapter creative activity, such as a crossword or cryptogram, related to topics discussed in the chapter
- Some personal life lessons I have learned throughout my life, directly connected to the topics of the chapter

Based on your personal interests and attitudes, you may decide to read all the parts or skip what you are not interested in, without losing out on learning.

Conventions Used in This Book

The following typographical conventions are used in this book:

Italic

Indicates new terms, URLs, email addresses, filenames, and file extensions.

`Constant width`

Used for program listings, as well as within paragraphs to refer to program elements such as variable or function names, databases, data types, environment variables, statements, and keywords.

> **TIP**
>
> This element signifies a tip or suggestion.

> **NOTE**
>
> This element signifies a general note.

> **WARNING**
>
> This element indicates a warning or caution.

Using Code Examples

Supplemental material (code examples, exercises, etc.) is available for download at *https://github.com/alod83/Learning-Generative-AI-Tools-for-Excel*.

If you have a technical question or a problem using the code examples, please send email to *support@oreilly.com*.

This book is here to help you get your job done. In general, if example code is offered with this book, you may use it in your programs and documentation. You do not need to contact us for permission unless you're reproducing a significant portion of the code. For example, writing a program that uses several chunks of code from this book does not require permission. Selling or distributing a CD-ROM of examples from O'Reilly books does require permission. Answering a question by citing this book and quoting example code does not require permission. Incorporating a significant amount of example code from this book into your product's documentation does require permission.

We appreciate, but do not require, attribution. An attribution usually includes the title, author, publisher, and ISBN. For example: "*Learning Generative AI Tools for Excel* by Angelica Lo Duca (O'Reilly). Copyright 2026 Angelica Lo Duca, 978-1-098-19017-0."

If you feel your use of code examples falls outside fair use or the permission given above, feel free to contact us at *permissions@oreilly.com*.

O’Reilly Online Learning

> **NOTE**
>
> For more than 40 years, O’Reilly Media has provided technology and business training, knowledge, and insight to help companies succeed.

Our unique network of experts and innovators share their knowledge and expertise through books, articles, and our online learning platform. O’Reilly’s online learning platform gives you on-demand access to live training courses, in-depth learning paths, interactive coding environments, and a vast collection of text and video from O’Reilly and 200+ other publishers. For more information, visit *https://oreilly.com*.

How to Contact Us

Please address comments and questions concerning this book to the publisher:

O’Reilly Media, Inc.

141 Stony Circle, Suite 195

Santa Rosa, CA 95401

800-889-8969 (in the United States or Canada)

707-827-7019 (international or local)

707-829-0104 (fax)

support@oreilly.com

https://oreilly.com/about/contact.html

We have a web page for this book, where we list errata and any additional information. You can access this page at *https://oreil.ly/learn-gen-ai-excel-1e*.

For news and information about our books and courses, visit *https://oreilly.com*.

Find us on LinkedIn: *https://linkedin.com/company/oreilly*.

Watch us on YouTube: *https://youtube.com/oreillymedia*.

Acknowledgments

First and foremost, I would like to thank my husband, Andrea, for his endless patience in listening to my thoughts, ideas, and sometimes long monologues about Excel and AI. Thank you, my love, for your words of comfort and for believing in me more than anyone else ever has. Thank you for the countless conversations, for your constant support, and for everything you do every single day.

My deepest gratitude goes to my two precious treasures, Giulia and Antonio. Through the innocence of childhood, you help me every day see small things as if they were the greatest wonders, appreciate every detail, and bring life even to inanimate objects, whether they are toys, pieces of fruit, or simple sheets of paper. Without you, the world would be a much sadder place. Thank you for awakening the child that still lives within me every day.

A special thank you goes to my acquisitions editor, Michelle Smith, for believing in this book and for helping bring it into the world. I am profoundly grateful to Corbin Collins, the book's development editor. Thank you, Corbin, for the many conversations, for your encouragement, and for your constant trust in this project. Thank you for believing in my unconventional ideas, from including crosswords and playful activities to supporting the idea of weaving a short novel into a technical book. Thank you for your long-term vision and for helping shape this work into what it has become.

I would also like to thank Dawn Griffiths, George Mount, and Sheldon Rodrigues for serving as technical reviewers. I consider each of you to be among the foremost Excel experts, and I am truly honored that you chose to contribute to this project. Without your insightful comments and thoughtful feedback, this book would not be even half of what it is today.

Additional thanks to the production team at O'Reilly: Ashley Stussy, Kate Dullea, Tonya Trybula, Ellen Troutman-Zaig, and Tim Stewart for keeping the many moving pieces aligned from manuscript to final release.

Finally, thank you to you, the reader, for choosing this book and for placing your trust in me. I truly hope this journey meets your expectations and perhaps even exceeds them.

Part I. Using ChatGPT in Excel

In Part I, you'll learn the basic mentality to combine Excel and AI. We'll focus on ChatGPT, but you can also apply the described methodology to other AI tools, such as Gemini and Claude, or even to conversational models you can install locally.

Chapter 1 introduces the basic concepts of Excel and explains why spreadsheets are a natural fit for AI assistance. It describes the benefits AI brings to Excel, including speed, accessibility, and support for complex analysis, and discusses situations where AI adds little value or can obscure simple logic. The chapter also walks through a simple example showing how AI can perform common Excel operations more quickly. It concludes with a survey of the main AI tools available for Excel, setting the stage for practical use in later chapters.

Chapter 2 introduces a practical framework for structuring conversations with ChatGPT using a narrative arc: context, main point, and action. This approach encourages a way of thinking that extends beyond any single AI tool. Rather than focusing on how to operate a specific system, the chapter helps you develop a reusable mindset for interacting with AI through well-structured prompts. The framework is illustrated using ChatGPT and applied to common Excel tasks. The chapter concludes with hands-on prompts for creating basic functions and formulas.

Chapter 3 explains how to use ChatGPT to accelerate common Excel tasks such as template creation, Autofill, and data enrichment. Building on the prompt structure introduced in Chapter 2, the chapter shows how to delegate structured Excel operations to ChatGPT using clear, repeatable instructions. Through practical examples and exercises, it highlights both the efficiency gains and the importance of validating AI-generated results.

Chapter 4 explores advanced Excel operations using ChatGPT, focusing on data analysis, PivotTables, and automation with macros and VBA. The chapter shows how ChatGPT can help identify relevant questions in a dataset, generate formulas to answer them, and guide the implementation of analytical workflows. It also demonstrates how to use ChatGPT to write, generalize, and debug VBA macros, enabling more complex automation while keeping the user in control of validation and final decisions.

The first appendix in Part IV takes the ideas from Part I and brings them to life. Instead of revisiting theory, it introduces a set of activities built around what I call *applied fiction*, an immersive learning format that borrows elements from a novel to guide you through realistic situations. Concepts emerge as the narrative progresses, allowing you to learn by following the story.

Top 10 New Articles:

- **Web 2.0 Submission Sites**
- **Is AI ML certification worth it**
- **What is Adobe Premiere Pro and How Its Works**
- **What is AI SEO**
- **Is Microsoft Power BI certification free**
- **What is an UX UI design certificate**
- **Is Adobe InDesign easy to use**
- **What is the use of Adobe Illustrator**
- **What is Adobe Photoshop in simple words**
- **Google Core Update 2026**

Chapter 1. Breaking the Ice: Introducing AI for Excel

If you have worked with Microsoft Excel, you've probably noticed how tiring and sometimes boring it is to perform repetitive operations such as formatting cells or doing calculations. By searching the web, you have probably found ways to navigate the jungle of data analysis, formulas, and formatting, but you keep encountering new problems that slow down your work. If you're *not* that familiar with Excel, you may be intimidated by the challenges you'll undoubtedly face when manipulating data. That has happened to me many times.

When I was 15, I started using Excel to catalog the books on my shelf. At the time, a quarter of a century ago, I was using Microsoft Excel 95. But later, even after I had been working with Excel for a long time, every time a new problem arose, I had to search the web or look at books for answers.

Today, everything has changed with the advent of generative AI, which has introduced new ways of working and speeding up tedious tasks. Generative AI is a potent tool that allows you to quickly manipulate and generate data. However, to make the best use of it, you need to know it, learn its limits, and see how far it can help you.

This chapter begins your journey to discovering how to integrate generative AI into Excel. This technology is evolving so fast that the AI models coming out today may seem old quickly. However, the *principles* described in this book are always valid, and you can apply them even if the technologies described become obsolete.

So what are we waiting for? Let's go. This chapter covers the following:

- Basic Excel concepts
- Why Excel needs AI
- The benefits of using AI in Excel
- When AI is not helpful in Excel
- Browsing the different AI tools for Excel

Basic Excel Concepts

As you probably know, Microsoft Excel is a very popular tool for processing different types of data. You can use it to display and manipulate tables, build and enrich graphs, extract knowledge, and much more. Whatever reason you use Excel, you will surely deal with data. *Data* is the raw element representing an observation or measurement before any interpretation or processing. Data can be numbers or text. For example, if the number of clouds you see in the sky today is 10, it means that 10 is a data point that represents the number of clouds. Today's date is another data point, and the category *cloud* is another data point. Imagine that every day, you count not only the number of clouds you see in the sky but also the number of birds, which represents a second category. You then organize everything into a table like Table 1-1.

Table 1-1. Sky dataset with the number of clouds and the number of birds organized by day

Date	Number	Category
2024-09-28	10	Cloud
2024-09-29	12	Cloud
2024-09-30	0	Cloud
2024-09-28	13	Bird
2024-09-29	7	Bird
2024-09-30	9	Bird

What you have there is a dataset. A *dataset* is, very simply, a set of data. In Excel, you can load several datasets, one at a time or all at once.

Before we move forward, let's clarify some Excel definitions you will see throughout the book (see Figure 1-1):

Cell

The basic unit of a table, located at the intersection of a row and a column of the table, where you can enter data or formulas.

Worksheet

A sheet within a workbook, organized into rows and columns, where data can be inserted, viewed, and manipulated.

Workbook

A file that contains one or more worksheets. It allows you to organize and manage different data tables within the same document.

Range

A set of selected cells. One *row* is a single horizontal row of cells, while a *column* is a vertical column of cells within a worksheet.

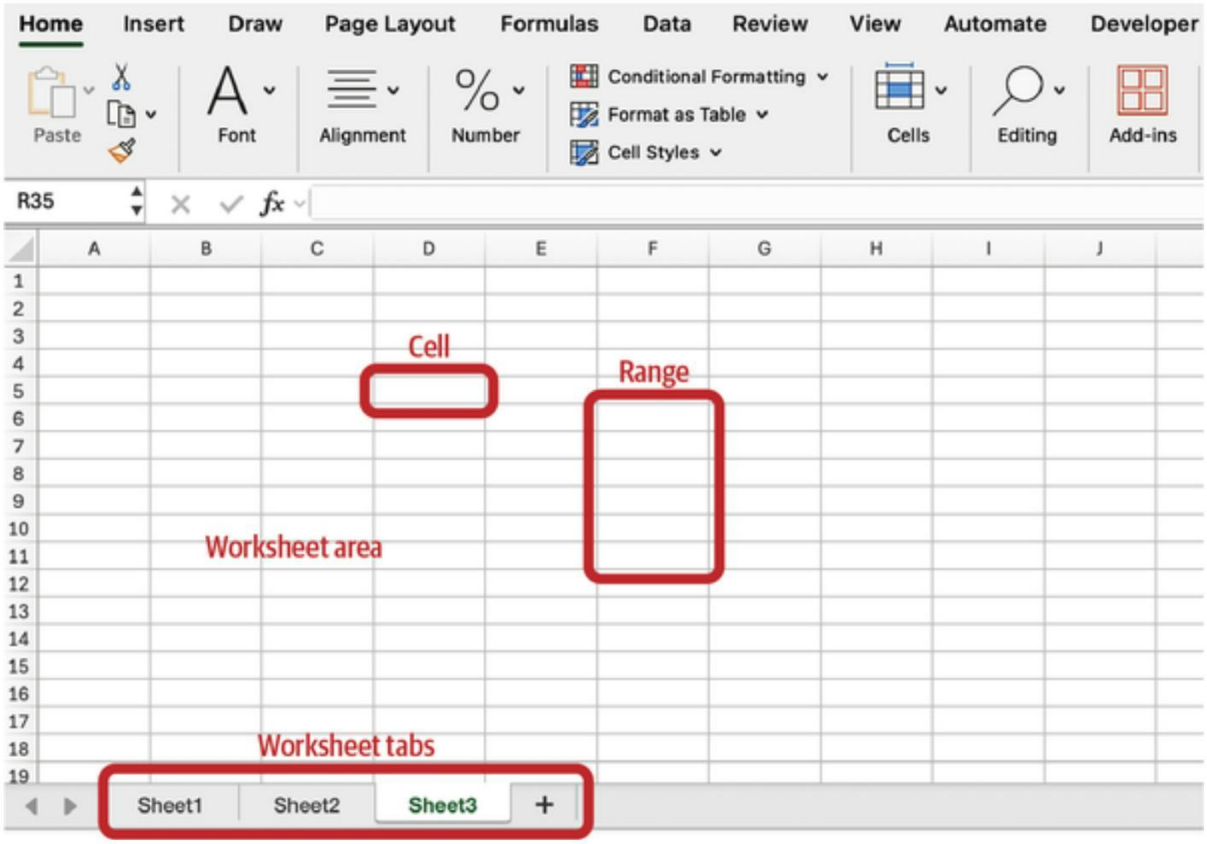


Figure 1-1. A visual representation of a cell, worksheet, multiple tabs within a workbook, and range

> **NOTE**
>
> To illustrate AI in Excel, much of this chapter will have you doing an exercise in Excel. If you know Excel like the back of your hand, this exercise will be child's play.

Consider the dataset on submarine cables in the United States released by Marine Cadastre.[1] The dataset is constantly updated. We will use the September 26, 2024, version, which you can download from the book's GitHub repository under *01/ch01.csv*.

This dataset contains the approximate location of commercial and research submarine cables in US waters. Most of these cables are used for communications, while a smaller portion is used for power transmission. In coastal areas, cables are usually buried under the seabed, while in the open sea, they are laid directly on the seabed. Just to get an idea of what we're talking about, Figure 1-2 shows a portion of the data plotted on a map of the US East Coast. Each line corresponds to a submarine cable.

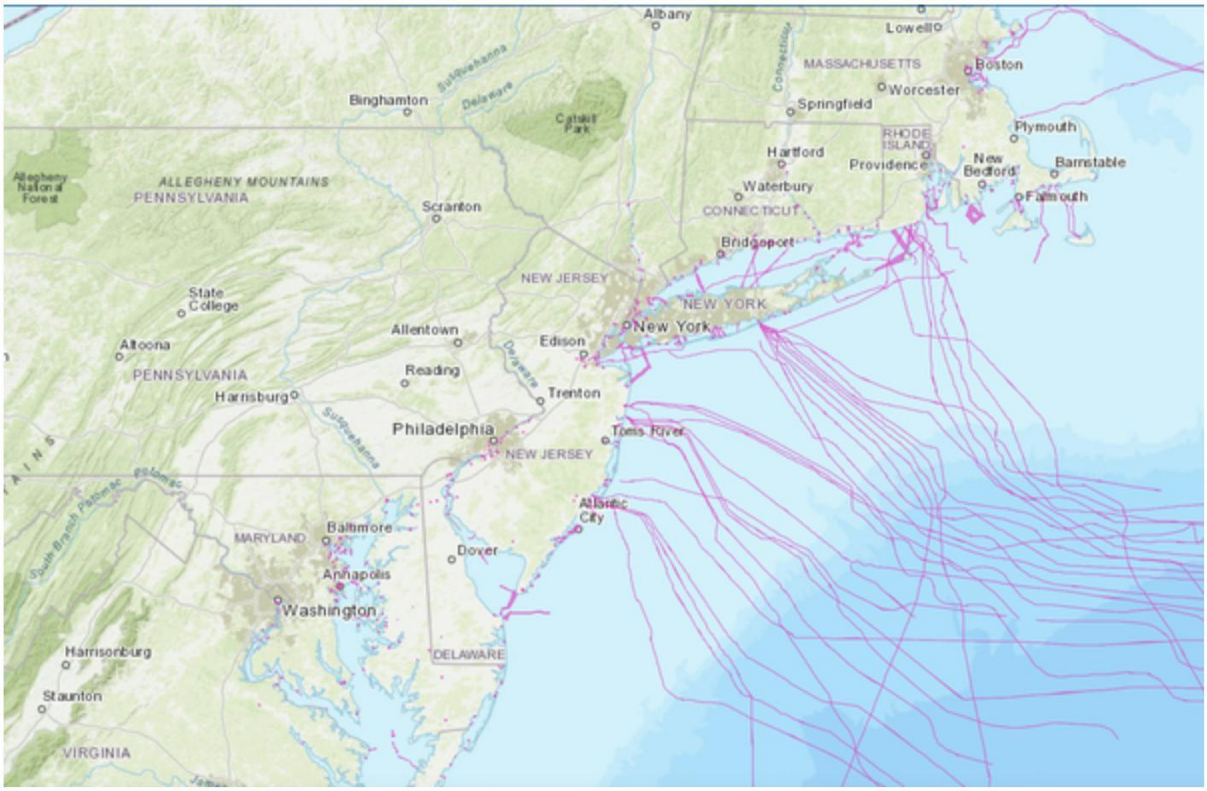


Figure 1-2. A map representation of a portion of the US submarine cable dataset

All the data shown in Figure 1-2 is organized in a table. Table 1-2 shows a portion of the US submarine cable dataset.

Table 1-2. A portion of the US submarine cable dataset

owner	cablesystem	region	SHAPE__Length	shortnam
AT&T Corp	Americas II	East Coast	108697.599067727	Americas I
GCI Communication Corp	AU-Aleutian Submarine Cable System	Pacific Northwest	1129434.70407587	AU-Aleutia
GU Holdings Inc	Curie	West Coast	1274007.14615703	Curie 1.1
Eversource Energy	Eversource HEEC Cable	East Coast	10111.1441552573	Eversource
GU Holdings Inc	Firmina	East Coast	1006081.05355044	Firmina
GU Holdings Inc	Grace Hopper	East Coast	1493252.13857506	Grace Hopp

The dataset contains the following fields:

`owner`

: The entity that owns the cable; this is a company or an organization

`cablesystem`

: The full name of the submarine cable system to which the cable belongs

`region`

: The geographic region where the cable is located

`SHAPE__Length`

The cable's length in meters

`shortname`

The short name of the submarine cable

`objectid`

A unique identifier for each object in the dataset

`SHAPE__Area`

The geographic area occupied by the cable in meters

`status`

The current status of the cable

In the remainder of this section, we'll use this dataset to see how to use Excel for different purposes, including data management, analysis, and visualization. You will see other cases later in the book.

Excel for Data Management

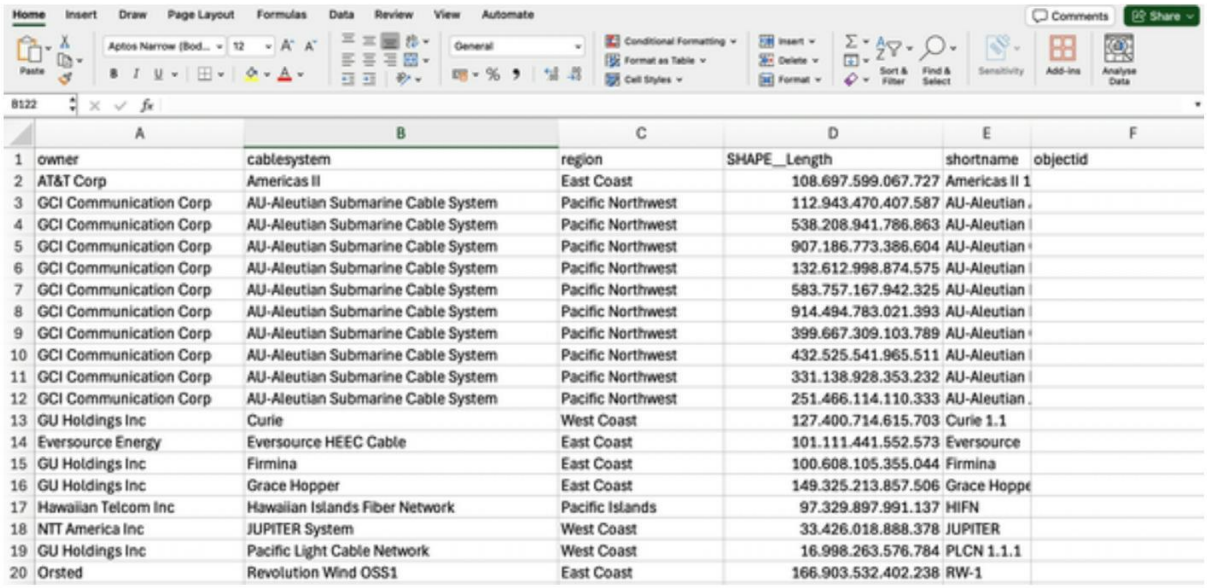
Data management involves collecting, organizing, storing, and maintaining data while ensuring its accessibility and security. The first step to managing data is to load the dataset as a table. Excel supports various formats, such as CSV, but to fully exploit Excel's potential, here you will save the dataset in XLSX format, which is Excel's proprietary format.

Download the dataset from the GitHub repository of the book under *01/ch01.csv* and save it in your local filesystem. Then, load it into Excel as follows:

1. Launch Excel, then select File → Open → Select your file. (In Excel 365 for Windows only, change the file type to All Files.)
2. Next, click Open → Choose the dataset from your filesystem → Check Delimited.
3. Click Next → Select Comma as a delimiter, then click Next.
4. Click Finish.

5. If a warning message appears about possible information loss, proceed to save the file as an Excel file by clicking Save As. On the screen that opens, select Excel Workbook or XLSX as the file format. Then click Save.

Figure 1-3 shows a portion of the dataset loaded into Excel.



	A	B	C	D	E	F
1	owner	cablesystem	region	SHAPE__Length	shortname	objectid
2	AT&T Corp	Americas II	East Coast	108.697.599.067.727	Americas II 1	
3	GCI Communication Corp	AU-Aleutian Submarine Cable System	Pacific Northwest	112.943.470.407.587	AU-Aleutian	
4	GCI Communication Corp	AU-Aleutian Submarine Cable System	Pacific Northwest	538.208.941.786.863	AU-Aleutian	
5	GCI Communication Corp	AU-Aleutian Submarine Cable System	Pacific Northwest	907.186.773.386.604	AU-Aleutian	
6	GCI Communication Corp	AU-Aleutian Submarine Cable System	Pacific Northwest	132.612.998.874.575	AU-Aleutian	
7	GCI Communication Corp	AU-Aleutian Submarine Cable System	Pacific Northwest	583.757.167.942.325	AU-Aleutian	
8	GCI Communication Corp	AU-Aleutian Submarine Cable System	Pacific Northwest	914.494.783.021.393	AU-Aleutian	
9	GCI Communication Corp	AU-Aleutian Submarine Cable System	Pacific Northwest	399.667.309.103.789	AU-Aleutian	
10	GCI Communication Corp	AU-Aleutian Submarine Cable System	Pacific Northwest	432.525.541.965.511	AU-Aleutian	
11	GCI Communication Corp	AU-Aleutian Submarine Cable System	Pacific Northwest	331.138.928.353.232	AU-Aleutian	
12	GCI Communication Corp	AU-Aleutian Submarine Cable System	Pacific Northwest	251.466.114.110.333	AU-Aleutian	
13	GU Holdings Inc	Curie	West Coast	127.400.714.615.703	Curie 1.1	
14	Eversource Energy	Eversource HEEC Cable	East Coast	101.111.441.552.573	Eversource	
15	GU Holdings Inc	Firmina	East Coast	100.608.105.355.044	Firmina	
16	GU Holdings Inc	Grace Hopper	East Coast	149.325.213.857.506	Grace Hoppe	
17	Hawaiian Telcom Inc	Hawaiian Islands Fiber Network	Pacific Islands	97.329.897.991.137	HIFN	
18	NTT America Inc	JUPITER System	West Coast	33.426.018.888.378	JUPITER	
19	GU Holdings Inc	Pacific Light Cable Network	West Coast	16.998.263.576.784	PLCN 1.1.1	
20	Orsted	Revolution Wind OSS1	East Coast	166.903.532.402.238	RW-1	

Figure 1-3. A portion of the dataset loaded into Excel

The data is held in a cell range and you want to convert it to a table. A *table* in Excel is a tool that helps you organize data more efficiently than a regular range of cells. Using a table has several benefits, such as using column names instead of cell ranges. It also makes it easier to work with charts and PivotTables.

Select a cell in the data, and then select Format as Table from the Ribbon Home tab to turn your raw data into a table. In the next window, select the table layout and leave the columns selected by default. Also, make sure there is a checkmark on the entry "My table has headers." Before we start working, let's give the table a name. Select the Table tab from the Ribbon and change the table's name to `MainTable` in the Table Name box on the left, as shown in Figure 1-4.

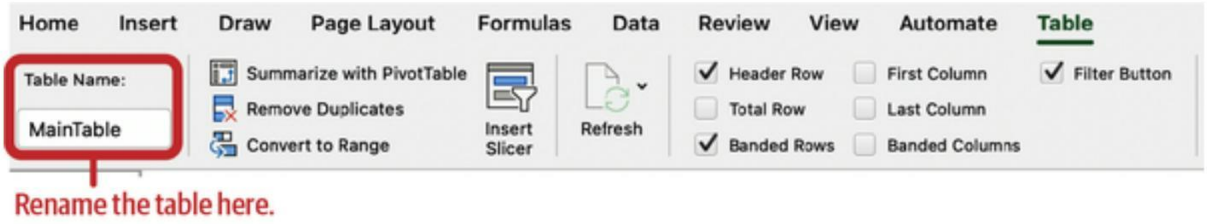


Figure 1-4. How to rename a table in Excel

At this point, you are ready to work.

> **NOTE**
>
> The Ribbon in Excel is the toolbar at the top of the Excel window. It organizes commands and features into tabs, such as Home, Insert, and Formulas, as shown in Figure 1-5.

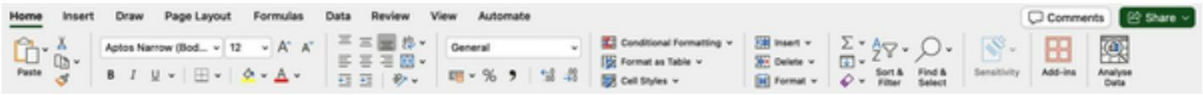


Figure 1-5. The Ribbon in Excel

For a long time, I ignored the existence of tables in Excel, working completely by hand. "Working by hand" meant exactly that: going row by row to apply the colors of my template, the formulas, and so on. Then, one fine day, I discovered tables, and I understood their importance when I realized that adding or removing rows to the table was very simple: everything updated automatically. If you haven't tried tables yet, I suggest you do so soon and you'll see that your Excel life will improve, as mine has.

Excel for Data Analysis

The second main purpose of using Excel is *data analysis*, which allows you to carry out mathematical and financial calculations. Excel offers various formulas and functions for performing numerical, financial, and statistical calculations, such as sums, averages, and so on.

Let's query the dataset by building a new table with the number of In Service, Planned, and Out of Service cables. We'll call it `StatusTable`. Many data points are unavailable, so they will be excluded from our calculation. To build the `StatusTable`:

1. Click in any cell outside the main table, such as cell J6, as shown in Figure 1-6.
2. Write the data values shown in Table 1-3, one in each cell, as shown in Figure 1-6.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 1-3. The data values for the StatusTable

status	count
In Service	
Planned	
Out of Service	

Figure 1-6. A possible place to insert data values

3. Format as a table by selecting all the cells involved and clicking on Format as Table, as described previously. Then, rename the table to `StatusTable` (Figure 1-7).

Figure 1-7. The `StatusTable` with the possible states

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

As a result, you have built a table with the following values:

The first column contains the status with the three values (In Service, Planned, and Out of Service).

The second column contains only the header (named `count`), which will include the number of items in the `MainTable` for each status.

To populate the count column values automatically, you must calculate the number of cables in the In Service state using Excel formulas. To access a formula, select it from the Formulas tab of the Ribbon or proceed as follows:

1. Double-click the first cell of the count column and insert the equals (=) symbol. This opens formula mode, where you can enter any formula you like.
2. You will use a *conditional* count, which counts rows from the status column of the main table only if they satisfy a criterion. Enter the following formula and press Enter:

```
=COUNTIF(MainTable[status], [@status])
```

The formula means that we use the `COUNTIF` function to count the cells in the status column of the `MainTable` only if they meet a particular criterion specified by the function's second parameter. The criterion is this: a cell is counted only if its value matches the current row's value in the `status` column. In other words, the `[@status]` function refers to the value of the `status` column in the currently processed row within the `StatusTable`. So, the formula is counting how many times the value of the current row's status column in the `StatusTable` appears in the status column of the `MainTable`.

NOTE

Excel uses a semicolon (`;`) as an argument separator in some countries, and in other countries, it uses a comma (`,`). In addition, Excel uses different names for functions in different languages, so in French, `COUNTIF` is `NB.SI`. Use Excel-Translator to translate a formula from English to another language.

When you press Enter, Excel populates the table with the results for each row. If it doesn't do this in your version of Excel, click the cell where you wrote the formula. A small square appears at the bottom right of the cell. Drag the square down to the last row of the `StatusTable`. The fields will be automatically populated with the correct values. Figure 1-8 shows the final result.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 1-8. The +StatusTable+ filled with missing values

You can achieve the same result as the previous example using PivotTables. A *PivotTable* is a table for dynamically summarizing, analyzing, exploring, and presenting data. Follow these steps:

1. Select the Insert tab from the Ribbon and then select PivotTable.
2. In the open window, select the `status` column in the `MainTable` (click the letter H to select the entire column), leave the New Worksheet option selected for the location, and then click OK. A new worksheet is created with the PivotTable.
3. In the box that opens on the right, select the `status` column and drag it into the ∑ Values area. In addition, drag the `status` column to the Rows section. Excel automatically populates the table, adding the states we had not considered in the previous example.

Figure 1-9 shows the newly built PivotTable.

Figure 1-9. The PivotTable for the different statuses

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Excel for Data Visualization

A third use of Excel is *data visualization*, which involves representing data through graphs, such as lines, bar graphs, or pie charts. Follow these steps to draw a bar chart of the PivotTable:

1. Click anywhere inside your PivotTable.
2. Select the Insert tab on the Excel Ribbon.
3. In the Charts group, open the Insert Column or Bar Chart group and select Bar Chart → Clustered Bar from the 2-D Bar section.

Figure 1-10 shows the resulting chart.

WARNING

In some cases, when you try to create a chart in a PivotTable, you may encounter an error that asks you to copy the data and then create the chart.

This simple exercise has demonstrated some of the main features provided by Excel:

Importing data in CSV format and converting it to the XLSX format

Formatting the data as a table

Naming the table

Calculating a simple formula

Using a PivotTable

Creating a chart in Excel

Excel can also be used for other purposes, such as automating processes and creating financial models. Throughout the book, we will explore these aspects and see how to best use AI to manage them.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 1-10. The bar chart built from the PivotTable

Exercise

Solve the following exercise using the submarine cable dataset:

1. Create a new worksheet by clicking the + symbol at the bottom of your page.
2. Rename the worksheet as PivotTable, using the right-click menu to select Rename.
3. Create a PivotTable showing the number of cables, organized by owner, for each status.

The solution to the exercise is in the book's GitHub repository, under *01/ch01_solution.xlsx*. The PivotTable uses the Count of `objectid`. The default value is Sum. To change it to Count, follow these steps:

1. Right-click the `objectid` in the PivotTable ∑ Values area.
2. Select Value Fields Settings → Count → OK.

Why Does Excel Need AI?

So far you have seen that Excel can be used for three different purposes: data management, data analysis, and data visualization. Indeed, there are other possible uses

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

of Excel as well. Figure 1-11 illustrates some of the many common uses of Excel.

Figure 1-11. Some of the ways to use Excel

Here's a bit more about those uses of Excel:

- Data management involves using Excel tables to store and organize data, making it easier to sort and filter. For example, you can create a table with customer information and use filters to organize the data.
- Data analysis involves extracting insights from a dataset, for example, applying and analyzing pivot tables.
- Data visualization involves creating charts, such as trends in sales data over time.
- Process automation automates repetitive tasks, such as using macros to format and update tables.
- Financial modeling creates a financial forecast, for example, to predict future cash flows.

Excel is already a complete data exploration and analysis tool, so extending it with external tools such as AI might seem excessive. However, there are at least two good reasons to integrate AI tools into Excel.

The first reason is the increase in *productivity*, the measure of the efficiency with which you carry out activities. Greater productivity means obtaining more significant results with less consumption of resources and time. Combining Excel and AI helps you do the following:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

- Carry out simple and complex tasks or suggest the steps to carry them out if you are new to Excel
- Automate repetitive tasks, such as data cleansing, formatting, or applying formulas, to save time and effort
- Process large quantities of data to extract insights
- Suggest appropriate visualizations for the data

The second reason to integrate AI into Excel concerns improving the *accuracy* of the results—that is, the ability to produce correct and precise results without errors:

- Reduce human errors, which can arise when you enter inputs or calculate values manually
- Recognize anomalies that could go unnoticed by the human eye

Let's look at an example to see how to combine AI and Excel to improve productivity. Consider the dataset from the previous exercise in which you calculated the number of occurrences for each cable status. Let's do the same exercise again using ChatGPT, one of the most popular AI assistants developed by OpenAI. The concepts described for ChatGPT can also be applied to other AI tools, such as Claude or Gemini. Throughout the book, you'll see other tools, including Copilot in Part II, and other add-ins, Python, and OpenAI in Part III.

First, start ChatGPT. If you don't have an account yet, visit *https://chatgpt.com* and proceed with registration if prompted. In this book, I use the Plus version of ChatGPT, which requires a monthly subscription, but the suggestions given also apply to the free version, even if the results may be slightly worse. Formulating correct and precise instructions is essential to using ChatGPT with Excel correctly. We will see the structure of various instructions later in the book.

To interact with ChatGPT, you use instructions called prompts. A *prompt* is text that describes what ChatGPT should do, and *prompt engineering* is the art and science of defining appropriate and effective prompts. There are several techniques for defining prompts:

Zero-shot prompting

> Give ChatGPT a task without providing examples. Use this technique if you want a quick answer. For example, write: "What is Microsoft Excel?"

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

One-shot prompting

Provide a specific example to tell ChatGPT how to respond. For example, write: “Microsoft Word is a word processing program developed by Microsoft, which allows you to create, edit, and format text documents professionally. Based on the previous definition, give a definition of Microsoft Excel.”

Few-shot prompting

Give the model multiple examples of how to complete a task. This helps the model better understand the context and format of the desired response. For example, you might provide several example definitions and then ask for the Microsoft Excel definition.

Instruction-based prompting

Give explicit instructions, often in natural language, to clarify the task. For example, you might say, “Explain what Microsoft Excel is to a nine-year-old.”

Chain of Thought (CoT) prompting

Guide ChatGPT to think aloud, breaking down the reasoning step by step, before arriving at an answer. This helps ChatGPT solve complex tasks that require logical reasoning. For example, you could prompt: “To explain what Microsoft Excel is, first describe what spreadsheets are, then explain how Excel uses them to organize data, and finally give an example of how Excel might be used to make a budget.”

Dynamic prompts

Adapt the prompt based on ChatGPT’s previous answers. This is an interactive approach in which the prompt evolves as the conversation context changes. For example, if ChatGPT just explained what a spreadsheet is, you might build on that answer and ask, “Based on your explanation, how might a spreadsheet be used to analyze a company’s expenses?”

Figure 1-12 summarizes the techniques described.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 1-12. Different prompting techniques

We will see and use these various techniques throughout the book. For now, let's do a test to evaluate ChatGPT's potential.

Excel and AI for Data Management

In the menu at the top left, select the ChatGPT model that is best suited to perform the calculations (GPT-5.2 at the time of writing this book). This exercise gives the data management function to ChatGPT, so we load the dataset directly into ChatGPT.

Excel and AI for Data Analysis

There are two ways of using Excel and ChatGPT. The first consists of carrying out all the analyses in ChatGPT and exporting the final file. The second is to use ChatGPT as an assistant to suggest formulas or procedures for carrying out tasks. This book uses a mixed approach: in some cases, we will use ChatGPT to conduct the analysis, and in others, we will use it as an assistant. You can choose which approach to follow based on the situations that arise. When we talk about Copilot in the second part of the book, we will perform all the operations within Excel, given that Copilot is already integrated into Excel.

Let's use the first approach, carrying out all the analysis in ChatGPT Plus and then exporting the final file. Perform the following operations:

1. Load the dataset *01/ch01.xlsx* in ChatGPT, using the paper clip, as shown in Figure 1-13.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 1-13. How to load a file into ChatGPT

2. Write the following prompt in the ChatGPT bar (after loading the dataset):

```
Build a pivot table for the column status and count the number of
occurrences for each category.
```

3. Press the Enter key. ChatGPT starts thinking. When it has completed processing, it displays descriptive text and a table similar to the one shown in Figure 1-14.

Figure 1-14. The pivot table built by ChatGPT to count different statuses

The pivot table calculated by ChatGPT is precisely the same as the one built by Excel in Figure 1-8, with a different formatting style.

The next step is to export the output as an Excel file. Write the following prompt in ChatGPT, still in the same chat conversation where you wrote the previous prompt:

```
Format the original table and the pivot table as two worksheets of a single
workbook in Excel. Calculate the pivot table using formulas instead of raw
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
data. Use column H of the original table in the pivot table formulas. Make
the output downloadable as an .xlsx file.
```

With this prompt, we ask ChatGPT to:

- Combine the original table and the generated pivot table in the same Excel file but under two different worksheets.
- Calculate the pivot table without hardcoded the numerical values into the table but applying the formulas to calculate them.
- Use the original table's H (status) column to perform calculations in the pivot table. This statement helps ChatGPT calculate the formula precisely.
- Make the file downloadable in Excel format.

As an output, ChatGPT generates a downloadable file. Download it and open it in Excel (Figure 1-15).

Figure 1-15. The Excel file generated by ChatGPT

The downloaded file is very similar to the one we produced manually, with some differences:

- The tables are not formatted like Excel tables.
- The pivot table is not a PivotTable as Excel usually calculates it, but it contains formulas to calculate the values correctly.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

In any case, it's not a bad starting point. In the same ChatGPT chat, write a new prompt, which is a modified version of the previous one, so that ChatGPT formats the data as Excel tables:

```
Format the original table and the pivot table as two Excel tables into two
worksheets of a single workbook in Excel. Calculate the pivot table using
formulas instead of raw data. Use column H of the original table in the
pivot table formulas. Make the output downloadable as an .xlsx file.
```

Download the generated file and open it with Excel. The data is now formatted as tables, as shown in Figure 1-16.

Figure 1-16. The Excel file generated by ChatGPT with formatted tables

The problem of the pivot table remains. Later in the book, you will see how to deal with a problem like this. For now, I hope I've convinced you that it's at least worth trying to use ChatGPT in Excel.

WARNING

The generated output may vary based on the ChatGPT model you are using. For example, using GPT-5.2, in my case, took 56 seconds to generate the file, but, at the end, it formatted the result as a table. In some cases, with GPT-4o, results were formatted as tables, but the PivotTable approximation didn't give the correct results. The `Count` column had `#REF!` for each row, and the formula for the first row was `=COUNTIF(#REF!, "In Service")`.

Excel and AI for Data Visualization

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Write the following prompt, still in the same ChatGPT chat:

```
Use Excel tools to build a bar chart for the pivot table. Add the chart as an
additional worksheet of the previous workbook. Make the output downloadable as
an .xlsx file.
```

Download the generated file. The results are very similar to those you obtained manually, as shown in Figure 1-17, with the only difference being the pivot table (which is not a classic Excel PivotTable).

Figure 1-17. The Excel file generated by ChatGPT with the bar chart

With this simple example, you have seen one of ChatGPT's two uses with Excel: automatic data analysis. The other type of use, which we will also see throughout the book, is as an assistant for suggesting formulas or procedures for data analysis.

If you have tried this exercise, I hope ChatGPT has not disappointed you, and that I have convinced you to continue reading this book. If you were faster than ChatGPT, it doesn't mean that ChatGPT isn't helpful to you, but rather that you already know Excel well,

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

and the integration with AI will only further improve your skills, as we'll see later in the book. For now, I promise you that using AI in Excel has benefits. Let's look at them.

The Benefits of Using AI in Excel

You may think that you need to use AI as your substitute in the sense that it automatically carries out tasks for you. In reality, in any field in which you use AI, including Excel, AI is always an assistant that can help you carry out your tasks more efficiently and accurately, but it can never wholly replace you. The last word will always be yours, and your task will be to ensure that the AI does not make gross errors, including hallucinations and biases, as we will see later in the book.

WARNING

AI *hallucinations* happen when the AI generates fake information. *Bias* is when the AI produces unfair or one-sided results.

In this book, I will talk about artificial intelligence, but not in the sense of an automatic process that completely replaces humans. We will not be getting into sci-fi robots that make autonomous decisions and ultimately destroy the universe. In this book, you will learn some strategies to use AI to augment *your* intelligence.

AI is nothing other than a significant part of human knowledge contained in a tool. Much of the human effort to understand how the world works is contained in a tool with impressive potential. This is fantastic and very powerful, but also risky. Therefore, the main benefit of using AI in Excel is not so much automating tasks and having the machine carry them out autonomously as it is to tap into a vast available knowledge base, which is the heritage of all humankind.

Your task is to learn to use this treasure in the best ways possible, and I will try to help you with that throughout the book. It is, therefore, not a question of acquiring concepts to automate more or less complex tasks, but instead a question of developing a mentality that will allow you to think with AI. You will always be the protagonist of the choices and analyses. AI will be your faithful travel companion at your service, helping you speed up tasks and be as precise as possible.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

EVALUATING AI RESULTS

To evaluate whether the AI is working well on your data, you can apply supervised learning of the AI without making errors. This involves testing the AI results on a subset of your data. Manually perform your calculations on some of the data using Excel and then repeat the same calculations using AI. Once you have consolidated the prompts for the AI and ensured that the AI responds correctly, let the AI perform the task on the entire dataset.

When AI Is Not Useful in Excel

AI is a potent tool. However, at least two cases indicate that its use is not yet mature (at least as of the writing of this book):

Large datasets

When working with large datasets, AI can encounter some difficulties regarding performance and accuracy. The analysis of large datasets requires many resources and computational capabilities. An AI model such as ChatGPT cannot provide these resources to ordinary users. For this reason, if you still want to use AI with large datasets, split the dataset into small parts that AI can process. For example, you could filter out unnecessary columns or rows from a dataset and provide only the data you want to analyze as input to AI.

Low-quality data

If you provide incomplete, noisy, or poorly structured data as input to AI, the quality of the results can negatively impact model performance, leading to incorrect predictions or distorted interpretations. AI will have difficulty providing valuable and accurate outputs if data is missing, duplicated, or contains systematic errors. This is an opportunity to repeat the classic maxim *garbage in, garbage out*, which means you can't extract a quality output from a bad input.

Despite these limitations, AI's potential continues to grow, and as technologies and infrastructure advance, improvements in these areas are expected.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Browsing the Different AI Tools for Excel

The introduction of AI into Excel has opened up new opportunities to simplify working with data, improve efficiency, and reduce human errors. In recent years, numerous add-ins have been developed that integrate AI directly into Excel, each with different characteristics to meet specific needs. Many of them may require a subscription. This section looks at some of the significant AI tools for Excel:

Numerous.ai

> Numerous.ai is an add-in that uses AI to make predictions, clean data, and generate new insights. It can summarize, categorize, classify, and cleanse data and text. Numerous.ai defines a specific function called `AI()`, which you can use to prompt ChatGPT directly in your worksheet.

GPT for Work

> GPT for Work is an add-in that uses AI for data cleaning and classification, text generation and summarization, and translations. It defines many functions that can be used directly in a cell, such as `GPT_EXTRACT()` and `GPT_CLASSIFY()`.

Power GPT

> Power GPT is an add-in that uses AI to answer questions from your data, decipher any formula with explanations in plain English, generate formulas, create SQL queries from simple descriptions, generate regular expressions, and so on.

AI-aided Formula Editor

> The AI-aided Formula Editor supports creating and optimizing complex formulas in Excel. It can also explain and format formulas. The command `write to cel` `l` enables you to copy the generated formula into a cell quickly.

Sally Suite

> Sally Suite is an add-in that uses AI to automate repetitive processes, create sheets, write functions, filter data, create or recommend

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

charts, translate sheets, perform data analysis, and even help generate Python code.

Ghostwriter

Ghostwriter is an add-in for generating sample data, translating sheets, writing prompts, generating formulas, and automating documentation. It is also available for Microsoft Word and PowerPoint, and can generate reports, write documents, create inline comments on your content, and more.

Shortcut.AI

Shortcut.AI is a platform that uses AI to automate and structure work in Excel. It can generate text, transform data, and apply reasoning across cells and workflows. Shortcut.AI defines reusable AI shortcuts that you can trigger from spreadsheets to turn prompts into consistent, repeatable results.

Excel Labs

Excel Labs is a collection of experimental features inside Excel used to test new capabilities. It provides the `LABS.GENERATIVEAI()` function to send prompts to an AI model.

The various add-ins are constantly updated, so new features that are not described in this book may become available. I always suggest you check each tool's official page and the terms and conditions before installing it.

Which tool(s) should you choose? It depends on your needs. This book does not use any specific add-ins (unless explicitly stated), but we will rely on ChatGPT in the first part and Copilot in the second part. I haven't talked much about Copilot so far because it is not an add-in but a tool integrated with Excel. We will see this later in the book.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

NOTE

I would like to end this chapter with a small sidenote. When I was little, like many children, I had an imaginary friend to whom I told all my facts and onto whom I projected all my childhood fears, doubts, and uncertainties. Unfortunately, since many years have passed since my childhood, I no longer remember the name of this imaginary friend. Today, however, I can say that I have a new friend, not exactly imaginary, but almost: ChatGPT (or AI in general), with whom I can interact in various ways. In this book, we'll see how to combine AI and Excel, but I've even read posts on social media from people who use AI as a therapist. Well, in this book, we will not use ChatGPT and AI as a therapist, but you can think of AI as your new imaginary friend, which unlike the one from your childhood, responds to your requests and is of great help to you. All you need to do is discover the key to open the door to this new imaginary friend.

Summary

Our journey together to discover the use of AI in Excel has just begun. In this chapter, you've seen why it makes sense to use AI in Excel and have gone through a brief example demonstrating its use:

- Integrating AI into Excel simplifies data management, analysis, visualization, and more tasks, improving overall efficiency.
- AI in Excel improves productivity and accuracy. AI allows you to perform tasks faster and more precisely. It also reduces human errors, ensuring more accurate and reliable results.
- Despite its many benefits, AI is not yet fully mature when processing massive datasets without advanced computational resources.
- Even the most powerful AI algorithm only works with quality data. Incomplete or noisy data can lead to misleading predictions or biased analyses.
- AI in Excel should be seen as an assistant to improve performance, not a complete replacement for human work. You should supervise the results generated by the AI to avoid errors.

In the next chapter, you will learn how to best use ChatGPT within your Excel sheets.

Activity

Solve the crossword puzzle in Figure 1-18. After completing it, you'll read the name of the AI tool you'll use in the next chapter. The solution is in Appendix D.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 1-18. Crossword puzzle to reveal the tool you will use next

Clues:

1. The basic unit of a table, located at the intersection of a row and a column of the table, where you can enter data or formulas
2. A sheet within a workbook, organized into rows and columns, where data can be inserted, viewed, and manipulated
3. An add-in for Excel providing the `AI()` function to prompt ChatGPT directly
4. An add-in for Excel providing the `GPT_EXTRACT()` function, which takes as input instructions and the data to process
5. A set of selected cells
6. The built-in AI tool provided by Excel. You'll use it in the second part of this book.
7. A section in the Excel Ribbon that groups related commands

[1] NOAA Office for Coastal Management, Submarine Cables dataset. This work incorporates U.S. Government material not subject to copyright protection under 17 U.S.C. 403.

Chapter 2. Getting Started with ChatGPT and Excel

In high school, I had to memorize trigonometry formulas by heart to solve math exercises. These were also very complex formulas that contained sines and cosines, additions and subtractions, and so on. After a lot of work, I managed to learn them well and was able to solve the trigonometry exercises perfectly. Periodically, however, I had to review the formulas aloud—otherwise, I would forget them. I was 17 at the time, and my brain was very trained. More than 20 years have passed since then, and I remember very little of those trigonometry formulas. Our brains tend to forget what we don't use, especially if it involves mnemonic formulas.

The same thing happens with all available Excel functions. It is impossible to remember all the functions that Excel makes available to carry out specific calculations. You may have a more or less general idea about the name of the function to use, but you often don't remember the syntax or the specific name. ChatGPT can help you with this, as it can tell you not only which function to apply each time but also how to apply it.

In this chapter, you'll see how to use ChatGPT to generate the formula or function you need. You will learn a strategy for communicating with ChatGPT to make it generate the correct output for your dataset: not something general but specific and valid to your dataset. When I talk about *communicating*, I'm referring to the literal meaning of transmitting a message to someone. The goal is to ensure that ChatGPT understands what you are talking about, so it can generate output as close as possible to what you want it to develop. So, you want to communicate well with ChatGPT. This communication strategy for ChatGPT involves organizing the conversation as a story, with its typical parts: the context, the main point, and the action. We will test this strategy for executing essential Excel functions, such as calculation, logic, and text functions.

In this chapter, you will learn more about:

ChatGPT for Excel

The structure of a conversation

Basic functions

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Let’s start with the first aspect: ChatGPT for Excel.

Introducing ChatGPT for Excel

There are two ways to combine ChatGPT and Excel, as shown in Figure 2-1.

Figure 2-1. The two possible uses of ChatGPT in Excel

In the first case, you can use ChatGPT as a *companion* or external tutor: ask specific questions about your dataset, and ChatGPT will help you answer them. Once you have the answers, apply them to your Excel file. This working method is helpful if you need clarification about how to write formulas or which functions to use, or if you need help answering the questions your data contains.

The second way to use ChatGPT is as a *direct editing tool*. In this case, you load the file directly into ChatGPT and work directly inside ChatGPT. In the end, you download the file with the changes already made by ChatGPT.

WARNING

In ChatGPT, file uploads are available only to authenticated users.

Whether you use ChatGPT as an assistant or as a direct editor, your job is to communicate with ChatGPT so that it understands what you want to talk about and helps you respond best. You can communicate with ChatGPT in two ways, as shown in Figure 2-2.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 2-2. The two ways of communicating with ChatGPT

A *single prompt* is a single input used to directly specify a request or question. It does not consider any previous context unless that context is made explicit in the prompt itself. For example: "Explain the formula for calculating the mean in Excel."

A *conversation* is a set of multiple prompts connected to each other. It is an interaction between the user and ChatGPT that considers the session's previous context, allowing for a continuous and coherent dialogue. Furthermore, it enables you to refine or expand a previously provided answer or ask subsequent questions about the ongoing conversation. For example, you might start with "Write the formula to calculate the average in Excel" and then ask, "Explain the formula."

WARNING

Remember that the suggestions provided by ChatGPT may not always be correct, so you should always check the suggestions or proceed iteratively to gradually achieve the desired result.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

The rest of this chapter focuses on using ChatGPT as an external assistant and working in conversation mode. However, everything I say about conversations also applies to prompts. At this point, we are ready to see the structure of a conversation (or a single prompt).

The Structure of a Conversation

Imagine that ChatGPT is a real conversation partner to whom you must say something. One important way humanity has communicated since ancient times is through stories. A *story* is defined primarily by a plot. You can organize your ChatGPT conversation as a simple story, where you are the narrator and ChatGPT is your audience, as shown in Figure 2-3.

Figure 2-3. How to organize a conversation with ChatGPT in the form of a story

Each story begins by describing the *context*, the set of all the information necessary to ensure that the story's audience (in your case, ChatGPT) understands what you are talking about. Then you present your *main point* to ChatGPT, such as calculating a formula or a specific function. Finally, you can ask ChatGPT to format the result in a certain way or to tell you why it used a specific formula.

Let's use a practical example to understand how to organize a conversation like a story. Suppose you have the Amazon Top 50 Bestselling Books 2009–2019 dataset, which is available on Kaggle under the CC0: Public Domain license. The dataset contains 550 books categorized into fiction and nonfiction. Table 2-1 shows an extract from the dataset.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 2-1. An extract from the Amazon Top 50 Bestselling Books 2009–2019 dataset

Name	Author	User Rating	Reviews	Price
10-Day Green Smoothie Cleanse	JJ Smith	4.7	17350	8
11/22/63: The Novel	Stephen King	4.6	2052	22
12 Rules for Life: An Antidote to Chaos	Jordan B. Peterson	4.7	18979	15
1984 (Signet Classics)	George Orwell	4.7	21424	6
5,000 Awesome Facts (About Everything!) (National Geographic Kids)	National Geographic Kids	4.8	7665	12

The dataset contains the following columns:

`Name`

: Name of the book

`Author`

: Author of the book

`User Rating`

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Amazon user rating: a number between 1 and 5

`Reviews`

Number of written reviews on Amazon

`Price`

The price of the book (as of 10/13/2020)

`Year`

The year(s) it ranked on the bestsellers list

`Genre`

Whether fiction or nonfiction

You can download the dataset directly from Kaggle in CSV format or from the book's GitHub repository in XLSX format under *02/books.xlsx*. In the latter case, the dataset is already formatted as an Excel table and named `books`.

Let's imagine we want to ask ChatGPT how to calculate the average price across all books. We can structure the prompt like a story: context, main point, and action. Let's start with the context.

Setting the Context

The *context* corresponds to a text that contains all the information ChatGPT needs to frame your dataset. The context defines the following:

1. The topic you are discussing
2. The structure of the dataset with column names
3. Guidelines on how to work in Excel, such as specifying the table name

As an alternative to points 2 and 3, you can directly upload the Excel file into ChatGPT. However, I suggest passing only the table structure and name if your file contains sensitive or large data.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

TIP

ChatGPT, by default, uses your input and conversations to train and improve the AI models ChatGPT is based on. You can disable data sharing and make your data private by following these steps:

1. On the ChatGPT web page, click on your profile icon at the bottom left and then select Settings from the menu.
2. In the Settings menu, look for the section "Data controls" and deselect the option "Improve the model for everyone."
3. Be sure to save any changes if required.

Turning this option off will prevent ChatGPT from using your data to improve models. However, you should still use caution when using ChatGPT to work with sensitive data.

In the case of the `books` dataset, organize the context into three parts. First, define the topic you are dealing with:

```
Consider a dataset relating to the Amazon Top 50 Bestselling Books 2009-2019.
```

Then specify the dataset structure with column names:

```
The dataset has the following columns:
Name - Name of the book
Author - The author of the book
User Rating - Amazon user rating: a number between 1 and 5
Reviews - Number of written reviews on Amazon
Price - The price of the book (as of 10/13/2020)
Year - The year(s) it ranked on the bestsellers list
Genre - Whether fiction or nonfiction
```

Finally, provide some guidelines on how to work with your dataset in Excel:

```
Suppose you import this dataset into Excel in a table named books and the column
names correspond to the dataset columns.
```

In total, the context part of the prompt should look like this:

```
Consider a dataset relating to the Amazon Top 50 Bestselling Books 2009-2019.
The dataset has the following columns:
Name - Name of the book
Author - The author of the book
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
User Rating - Amazon user rating: a number between 1 and 5
Reviews - Number of written reviews on Amazon
Price - The price of the book (as of 10/13/2020)
Year - The year(s) it ranked on the bestsellers list
Genre - Whether fiction or nonfiction
Suppose you import this dataset into Excel in a table named books and the column
names correspond to the dataset columns.
```

This prompt is the first message of a more extended conversation, or you can combine it with the rest of the story, which we'll see shortly, and build a single prompt. Go to the ChatGPT website and give it the previous context prompt as input. To open the ChatGPT page follow these steps:

1. Connect to *https://chatgpt.com*, log in, if prompted, and start a new conversation.
2. Copy the context prompt defined previously and check the result.

ChatGPT can respond with suggestions on what operations or types of analysis you can perform using this dataset. Out of pure curiosity, I fed this prompt to ChatGPT five times, starting a new chat each time, and I noticed that ChatGPT more or less always suggests doing the same operations, even if applied to different dataset columns. Operations include calculation of statistics, filtering and sorting operations, analysis, and visualizations of specific data. Also, in general, the answer can vary based on the type of subscription.

Once you've defined the context, you can specify the main point of the prompt or conversation. We'll explore this further in the next section.

Defining the Main Point

The story's *main point* is the problem that ChatGPT needs to solve, or more generally, the question you want to ask ChatGPT. The structure of the main prompt comprises four main elements:

The action

> This indicates the operation you want ChatGPT to perform. It's formulated clearly and directly. It includes verbs that describe what you want to achieve (e.g., *Write*, *Explain*, *Analyze*).

Objective

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

This describes your request's final result or purpose, that is, what you want to achieve through the action. It must be specific and linked to the problem to be solved or the task to be completed.

Additional details

These define additional information that has not been described in the context and should help ChatGPT to understand the request, such as available data, relevant parameters, or restrictions. This helps personalize the response by making it more appropriate.

The level of complexity

This specifies the necessary level of detail in the response, adapting it to the user experience. For example, it can be basic (for beginners) or advanced (for advanced users).

In some cases, as we will see shortly, you can omit some elements. For example, after defining the context as explained previously, use this main point prompt to write the formula in Excel to calculate the average price of books by genre:

Action: `Write`

Objective: `the formula in Excel to calculate the average price of books`

Additional details: `-`

Level of complexity: `-`

Here is a possible output produced by ChatGPT (plus some explanatory text):

```
=AVERAGE(books[Price])
Explanation:
books[Price]: Refers to the Price column in the table named books.
AVERAGE: Calculates the mean value of the numbers in the Price column.
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

As you can see, ChatGPT has generated a formula adapted to your dataset because it contains exactly the names of the `books` table and the columns (`Genre` and `Price`) that you previously set in the context.

NOTE

ChatGPT returns the formula in English by default. However, if your Excel is set in another language, you should also indicate the language in which Excel is used for ChatGPT. For example, if you use the Italian version of Excel, you can write the following prompt:

```
Write the formula for the Italian version of Excel to
calculate the average price of books.
```

In this case, the generated formula is: `=MEDIA(books[Price])`.

You can adapt the prompt structure based on your use of ChatGPT. Generally, you can structure a main point prompt differently based on the different uses of ChatGPT, as shown in Figure 2-4:

As an *assistant*, ChatGPT can guide you through using Excel, helping you understand how to create formulas, use functions, or customize spreadsheets to fit your needs. For example, you could ask ChatGPT: "Write the formula in Excel to calculate the average price of books by genre."

As a *teacher*, ChatGPT can clarify complex Excel concepts, such as macros, PivotTables, and data analysis, making them more understandable. For example, you could ask ChatGPT: "How do PivotTables work to analyze the total number of reviews for each year in the dataset?"

As an *analyst*, ChatGPT supports you in interpreting data, suggesting ways to visualize information through graphs, or identifying trends and patterns. For example, you might ask: "What are the five books with the highest average rating, considering only fiction books?"

For *troubleshooting*, ChatGPT can help you fix common problems in Excel, such as formula errors, formatting problems, or difficulties importing data. For example, you could ask: "What is the error in this [formula]?"

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 2-4. The roles ChatGPT can play in helping you answer your questions

Throughout the book, we will see how to define the various questions based on the task you must carry out in Excel. For now, it's enough to know that there are four types of main points.

The main prompt's contents change depending on the ChatGPT role, as described in Table 2-2. The table shows a structured guide for defining prompts for use with ChatGPT based on four main roles: assistant, teacher, analyst, and troubleshooting. For each role type, the table shows the structure of the main point prompt and an example of its use.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 2-2. The contents of the main prompt based on the ChatGPT role

Role	Action	Objective	Additional details	Level of complexity
Assistant	Clear and direct request. Verbs: Write, Create, Show, Calculate, Define	Get a formula or function to complete a specific task	Specific context, such as columns or functions	Not required
Teacher	Request a detailed explanation or step-by-step guide. Verbs: Explain, Show, Describe, Guide, Illustrate	Understand a complex concept or learn to use a function	Add a practical example or data relevant to the concept explained	Specify the level: beginner or advanced
Analyst	Request analysis or visualization suggestions. Verbs: Analyze, Suggest, Identify, Visualize	Extract meaningful trends, patterns, or insights from your data	Specific context, such as columns	Not required, but it may be helpful to specify the chart type

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Role	Action	Objective	Additional details	Level of complexity
Troubleshooting	Request assistance to resolve a specific problem. Verbs: Solve, Correct, Identify, Help, Explain	Request assistance to resolve a specific problem	Provide details such as the formula or context of the problem	Not required

Once you have defined the prompt's main point, you can proceed to its final part, which corresponds to the end of the story.

Setting the Action

The action of the story tells ChatGPT what to do after performing the main task. You have three main options. You can ask ChatGPT to do the following:

> Explain the steps that are required to generate the response in Excel. For example, if we ask ChatGPT to create a formula, we can now ask what the steps are to insert it into Excel.
>
> List alternative solutions.
>
> Calculate the answer to the question directly in an Excel file to export as output. You can only use this solution if you initially loaded the Excel file containing the entire dataset or a portion of it.

Returning to our example prompt ("Write the formula in Excel to calculate the average price of books"), we can ask ChatGPT to list the steps for entering the formula into

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Excel using the following prompt (always within the same conversation that contains the context and main point):

```
List the steps to include the proposed formula in Excel.
```

In my case, ChatGPT generates a detailed list of steps. Here, I report only the top-level steps:

```
Step 1: Open the Dataset
Step 2: Convert Data to a Table (Optional)
Step 3: Select a Cell for the Formula
Step 4: Enter the Formula
Step 5: Verify the Result
Optional: Format the Result
Step 6: Save the File
```

At this point, you have learned how to communicate with ChatGPT. Organize your conversation like a story: context, main point, and action. You can structure your chat as a single prompt to feed to ChatGPT or as three separate prompts within the same conversation: one for the context, one for the main point, and one for the action.

TIP

I suggest using three separate prompts in the same conversation so you can reuse the same context for multiple questions or main points.

So far, we have seen the generic structure of a prompt in ChatGPT. The more knowledge you have about the problem you want to solve, the better the results ChatGPT will produce. The next exercise looks at a practical example to see how to best structure prompts.

Exercise: Structuring Prompts

Consider again the Amazon Top 50 Bestselling Books 2009–2019 dataset. Start a new conversation with ChatGPT and write the following context prompt (defined previously):

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
Consider a dataset relating to the Amazon Top 50 Bestselling Books 2009-2019. The
dataset has the following columns:
Name - Name of the book
Author - The author of the book
User Rating - Amazon user rating: a number between 1 and 5
Reviews - Number of written reviews on Amazon
Price - The price of the book (as of 10/13/2020)
Year - The year(s) it ranked on the bestsellers list
Genre - Whether fiction or nonfiction.
Suppose you import this dataset into Excel in a table named books and the column
names correspond to the dataset columns.
```

Then write the main point prompt for each of these tasks:

1. *Task 1:* Calculate the highest price.
2. *Task 2:* Explain how to create a PivotTable to analyze the total sum of reviews for each year and sort the results in ascending or descending order.
3. *Task 3:* Find the error in this formula to calculate the average price of books with a user rating higher than 4.5: `=AVERAGEIF(books[User Rating]>4.5, books[Price]))`.

Let's proceed with the resolution of the exercise, starting with the first task.

Task 1

Use ChatGPT as an assistant and write the following prompt:

```
Write the formula in Excel to calculate the highest price of books. Use the Price
column from the books table.
```

ChatGPT should respond with an output like this:

```
To calculate the highest price of books in the Price column from the books table,
use the following formula in Excel:
=MAX(books[Price])
Explanation:
  • books is the name of the table.
  • [Price] refers to the column named Price within the books table.
  • MAX is the function that returns the highest value in the specified range.
Place this formula in any cell, and it will display the maximum price of the
books in the table.
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Paste the formula suggested by ChatGPT into the Excel sheet, as shown in Figure 2-5.

Figure 2-5. The result of Task 1

Task 2

Use ChatGPT as a teacher and write the following prompt:

```
Explain how to create a PivotTable in Excel to analyze the total sum of reviews
for each year and sort the results in ascending or descending order. Use the
Reviews column as the values and the Year column as rows. Use a base level
of complexity.
```

In my case, ChatGPT produced the following output, with a simple explanation of the steps to follow to create the required PivotTable:

```
Here is a step-by-step guide to creating a PivotTable in Excel to analyze the
total sum of reviews for each year, using the Reviews column as values and the
Year column as rows:

Step 1: Select Your Data
  1. Click anywhere inside the books table to ensure all columns (Year, Reviews,
     etc.) are recognized.
  2. Go to the Insert tab on the Ribbon and click PivotTable.

Step 2: Set Up the PivotTable
  1. In the Create PivotTable dialog:
    • Ensure the correct table or range (books) is selected.
    • Choose where to place the PivotTable:
      - Select New Worksheet (recommended) or Existing Worksheet.
    • Click OK.

Step 3: Add Fields to the PivotTable
  1. In the PivotTable Fields there are:
    • Drag the Year field to the Rows area. This will list all the years in the
        rows of the PivotTable.
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
    • Drag the Reviews field to the Values area. By default, Excel will sum the
      Reviews.

Step 4: Sort the Results
  1. Click the dropdown arrow in the Row Labels section of the PivotTable.
  2. Choose Sort by... and select either Smallest to Largest or Largest to
     Smallest to sort the total reviews by year in ascending or descending order.

Step 5: Optional - Format and Analyze
  1. If desired, format the numbers in the PivotTable:
    • Right-click any number in the Sum of Reviews column.
    • Select Number Format and choose a format like Number (with thousands
      separators) for clarity.
  2. Review the table to see the total reviews for each year sorted as specified.
```

If you follow the instructions provided by ChatGPT step-by-step, you should get the PivotTable shown in Figure 2-6.

Figure 2-6. The PivotTable produced as a result of Task 2

Task 3

Use ChatGPT to troubleshoot by writing the following prompt:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
Find and explain the error in the provided formula and suggest the correct
version to calculate the average price of books with a user rating higher
than 4.5: =AVERAGEIF(books[User Rating]>4.5, books[Price]))
```

ChatGPT responds by explaining that the formula contains several errors. After explaining the correct syntax of the `AVERAGEIF` function, it provides the correct formula:

```
=AVERAGEIF(books[User Rating], ">4.5", books[Price])
```

Paste the formula into an Excel sheet, as shown in Figure 2-7.

Figure 2-7. The result of Task 3

WARNING

If your computer uses a comma (,) as a decimal separator instead of a period (.), the provided formula may return an error or unexpected results. This is because Excel interprets numeric values differently depending on your system locale.

To solve the problem, you can change the general settings of your computer, or you can adapt the formula to the local decimal point. For example, if you use a comma as a decimal separator, write the formula as follows:

```
=AVERAGEIF(books[User Rating]; ">4,5"; books[Price])
```

Note that the period in the value 4.5 is replaced with the comma in the value 4,5, and the argument separator changes from “,” to “;”.

Challenge: What information must you add to the prompt context so that ChatGPT automatically generates the formula using your local decimal point?

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Now that we have consolidated the conversation structure with ChatGPT, let's examine the concept of formulas and functions in Excel and see how we can use ChatGPT to speed up their creation.

Basic Functions and Formulas

I have always been passionate about mathematics, particularly the study of functions and formulas. I remember with pleasure that when I was in high school, my mathematics teacher explicitly prevented me from answering the questions she asked the class about which function to use in a certain situation because she knew I already knew the answer. My passion for functions and formulas has spanned the years, and I have carried it to this day.

There is a fundamental difference between the concepts of function and formula:

> *Functions* are predefined operations that simplify complex or repetitive calculations, such as adding numbers, calculating averages, or searching for specific values (for example, `SUM`, `AVERAGE`).
>
> *Formulas* are user-defined expressions that perform calculations or manipulations on data. They always start with an equals sign (`=`) and can combine operators, constants, references, and functions (such as `=A1+A2`).

This section discusses the main types of functions with practical examples using the `books` dataset. You will also see how ChatGPT can help you use these functions to build basic formulas.

You can group the functions provided by Excel into the following six macro-categories:

Calculation

> Perform mathematical operations like average, sum, maximum, and minimum

Counting

> Count cells, values, or specific conditions

Logical

> Evaluate conditions or perform actions based on conditions

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Search

Find specific values or positions within data

Text manipulation

Modify or analyze text strings

Date and time

Handle operations involving dates and times

Let's examine each group of functions individually and consider how ChatGPT can help you define them faster based on your data.

Calculation

Use the *calculation functions* to perform basic mathematical and statistical operations, such as adding, averaging, rounding, or finding the maximum and minimum value in a set of data. Examples include:

`=SUM(A1:A10)`: Calculates the sum of the values in cells A1 to A10

`=AVERAGE(A1:A10)`: Calculates the average value of a range

`=MAX(A1:A10)`: Returns the maximum value in a range

`=ROUND(A1, 2)`: Rounds the value in A1 to two decimal places

You can use ChatGPT as an assistant to generate the right formula for your data. Start the conversation with ChatGPT by defining the context as described previously and then define the main point as follows:

Action: `Write`

Objective: `the formula in Excel to calculate [calculation function] [column].`

Additional details: `Use the table [table name].`

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Level of complexity: -

For example, if you want to calculate the total sum of all reviews in the `books` table, use the following prompt:

```
Write the formula in Excel to calculate the sum of Reviews. Use the table books.
```

ChatGPT responds by indicating the following formula:

```
=SUM(books[Reviews])
```

Exercise

After defining the context for the `books` table, write the prompt for the following tasks:

1. Calculate the average of the reviews in the `books` table.
2. Calculate the sum of the prices of all the books.
3. Calculate the average price of books in the “Fiction” genre.
4. Identify the highest price among books in the “Non Fiction” genre.
5. Calculate the minimum price of books in the “Non Fiction” genre.

Test every output produced by ChatGPT in Excel.

Counting

Use the *counting functions* to count cells based on certain criteria, such as counting non-empty cells, counting cells that meet a specific condition, or multiple criteria at once. Examples include:

`=COUNTA(A1:A10)`: Count all non-empty cells.

`=COUNTIF(A1:A10, ">5")`: Count cells with values greater than 5.

`=COUNTIFS(A1:A10, ">5", B1:B10, "Fiction")`: Count cells that match multiple criteria.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

You can use ChatGPT as an assistant to generate the right formula for your data. Start the conversation with ChatGPT by defining the context as described previously and then define the main point as follows:

Action: `Write`

Objective: `the formula in Excel to count cells [criterion].`

Additional details: `Use the table [table name].`

Level of complexity: `-`

For example, if you want to count all books with reviews greater than 4, write the following prompt:

```
Write the formula in Excel to count cells with review > 4. Use the table books.
```

ChatGPT generates the following output:

```
=COUNTIF(books[Reviews], ">4")
```

Exercise

After defining the context for the `books` table, write the prompt for the following tasks:

1. Determine how many books belong to the “Fiction” genre.
2. Count the number of books that have a price of 0.
3. Check how many books have received more than 1,000 reviews.
4. Count how many times the author “Mark Twain” appears in the `Author` column.
5. Count the total number of books in the table, regardless of specific values.

Test every output produced by ChatGPT in Excel.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Logical

Use the *logical functions* to perform conditional operations returning different results based on true or false conditions. Examples:

> `=IF(A1>10, "High", "Low")`: Returns `High` if A1 is greater than 10, otherwise `Low`
>
> `=AND(A1>10, B1<5)`: Returns `TRUE` if both conditions are satisfied
>
> `=OR(A1>10, B1<5)`: Returns `TRUE` if at least one of the conditions is satisfied

Search

Use the *search functions* to find specific values within a range of data. Examples:

> `=VLOOKUP(1001, A1:D10, 3, FALSE)`: Looks for the value `1001` in the first column of the range and returns the value of the third column
>
> `=INDEX(A1:A10, 5)`: Returns the value in the fifth row of the range
>
> `=MATCH(50, A1:A10, 0)`: Finds the position of the value `50` in the range

You can use ChatGPT as an assistant to generate the right formula for your data. Start the conversation with ChatGPT by defining the context as described previously and then define the main point as follows:

> Action: `Write`
>
> Objective: `the formula in Excel to search [criterion].`
>
> Additional details: `Use the table [table name].`
>
> Level of complexity: `-`

For example, if you want to find the genre of the book *10-Day Green Smoothie Cleanse*, write the following prompt:

```
Write the formula in Excel to search the Genre for the book “10-Day Green
Smoothie Cleanse”. Use the table books.
```

ChatGPT generates an output like this:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
=VLOOKUP("10-Day Green Smoothie Cleanse", books, 7, FALSE)
```

- Explanation:
- "10-Day Green Smoothie Cleanse": The value you are searching for in the first column of the books table.
- books: The name of the table.
- 7: The column index for Genre in the books table (assuming Genre is the 7th column).
- FALSE: Ensures an exact match.

Exercise

After defining the context for the `books` table, write the prompt for the following tasks:

1. Find the author of the book *The Hunger Games*.
2. Find the genre of the book with the maximum price.
3. Find the price of the book *The Great Gatsby*.
4. Find the year of the book with the maximum number of reviews.
5. Find the name of the book with the lowest price.

Test every output produced by ChatGPT in Excel.

Text Manipulation

Use *text formulas* to manipulate and manage text strings. They allow you to merge text, extract specific characters, and determine the length of a string, improving the processing of textual data. Examples include:

`=CONCAT(A1, B1)`: Combines the text in cells A1 and B1

`=LEFT(A1, 5)`: Extracts the first 5 characters of the text in A1

`=LEN(A1)`: Returns the length of the text in A1

You can use ChatGPT as an assistant to generate the right formula for your data. Start the conversation with ChatGPT by defining the context as described previously and then define the main point as follows:

Action: `Write`

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Objective: `the formula in Excel to [text manipulation] [involved columns].`

Additional details: `Use the table [table name].`

Level of complexity: `-`

For example, if you want to extract the first name of each author, write the following prompt:

```
Write the formula in Excel to extract each author's first name in the Author
column. Use the table books.
```

ChatGPT generates the following output:

```
=LEFT(books[@Author], FIND(" ", books[@Author]) - 1)
Explanation:
  • books[@Author]: Refers to the current row's value in the Author column of the
    books table.
  • FIND(" ", books[@Author]): Finds the position of the first space in the
    author's name.
  • LEFT(books[@Author], FIND(" ", books[@Author]) - 1): Extracts all characters
    from the start of the name up to (but not including) the first space, giving
    the first name.
```

Exercise

After defining the context for the `books` table, write the prompt for the following tasks:

1. Remove extra spaces from a title.
2. Extract the last name of the author.
3. Transform the title to lowercase.
4. Combine the book title and author.
5. Find the length of a book title.

Test every output produced by ChatGPT in Excel.

Date and Time

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Use *date* and *time* formulas to work with dates and times, allowing you to get the current date and time or extract specific parts of a date. Examples include:

`=TODAY()`: Returns the current date

`=YEAR(A1)`: Extracts the year from a date

`=DAYS(B1, A1)`: Calculates the difference between two dates in days

You can use ChatGPT as an assistant to generate the right formula for your data. Start the conversation with ChatGPT by defining the context as described previously and then define the main point as follows:

Action: `Write`

Objective: `the formula in Excel to calculate [date operation].`

Additional details: `Use the table [table name].`

Level of complexity: `-`

For example, if you want to extract the year from today's date, write the following prompt:

```
Write the formula in Excel to calculate the year from today's date. Use the table
books.
```

ChatGPT generates the following output:

```
=YEAR(TODAY())
```

Exercise

After defining the context for the `books` table, write the prompt for the following tasks:

1. Calculate the number of years that have passed since a book was published.
2. Find books published in the last five years.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

3. Calculate the difference in days between today and the year a book was published.
4. Calculate the age of a book in months.
5. Check if a book was published in a leap year.

Test every output produced by ChatGPT in Excel.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

THE FOREST FOR THE TREES

At the beginning of the chapter I told you about my passion for mathematics since I was a child. Over the years, to this passion I have added a passion for data and everything it represents. When you start loving data, you'll see it everywhere. Just a few days ago, during a car trip, I was sitting in the passenger seat, and my husband was driving. As the car rolled along the highway, I admired the passing landscape. At a certain point, something strange appeared: a hill with trees lined up and arranged in order. And I noticed how similar those trees were to the rows and columns of a table.

To immortalize the moment, I took the photo shown in Figure 2-8.

Figure 2-8. The photo I took during my trip

Out of curiosity, after arriving at my destination, I gave Figure 2-8 as input to ChatGPT and asked it to extract the number of rows and columns of trees, organize them in the form of a table, and format them as an Excel file. I used this prompt:

```
Consider the attached picture containing trees organized in rows and columns.
Build a table with as many rows and columns as the number of tree rows and
columns in the picture. Format the resulting table as an Excel Table
downloadable as an XLSX file.
```

The result was amazing. ChatGPT counted 20 tree columns and 10 tree rows and organized them into an Excel file, as shown in Figure 2-9.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 2-9. The table built by ChatGPT based on the picture shown in Figure 2-8.

Counting the rows and columns by hand, I ascertained that there are exactly 20 columns and about 10 rows. I say *about* because it's difficult to extract the exact number of rows. However, the result provided by ChatGPT was impressive.

From this example you can see the potential of ChatGPT, which can extract tables even from images. And it is precisely this, as well as other concepts, that we will deal with in the next chapter.

Summary

We have reached the end of this chapter, and you have now seen some basic strategies for using ChatGPT to generate appropriate prompts for your dataset:

ChatGPT can simplify complex tasks like generating functions, guiding data analysis, and resolving errors efficiently.

Prompts are organized as narratives with three key parts:

– *Context:* Define the dataset and requirements.

– *Main point:* Specify the task or question.

– *Action:* Outline the desired outcome or next steps.

ChatGPT can play four different roles while assisting you:

– *Assistant:* Generate formulas and solve specific tasks.

– *Teacher:* Explain complex concepts and guide step-by-step processes.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

- *Analyst:* Identify patterns, trends, and insights in data.
- *Troubleshooter:* Diagnose and correct formula errors or functionality issues.

Excel essential functions can be grouped into six categories: calculation, counting, logical, search, text manipulation, and date and time.

ChatGPT can help you to build customized formulas for your specific dataset.

In the next chapter, you will learn how to speed up common operations in ChatGPT.

Activity

Complete the following text using the keywords indicated below. The solution is in Appendix D.

ASSISTANT

CALCULATION

CHATGPT

CONTEXT

ACTION

MAIN POINT

PROMPT

VLOOKUP

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

In this chapter, you learned how to structure an effective _____ to best use _____ for Excel, organizing it into three fundamental parts: context, main point, and action. The _____ defines the data and information necessary to frame the problem, the _____ describes the action required, while the _____ indicates how to use the result.

ChatGPT can be used as a/an _____ to perform various tasks, such as generating formulas for _____ (for example, summing or averaging) or performing search operations with functions like _____. This combination of strategies helps you improve productivity, automate complex calculations, and quickly resolve problems in your datasets.

Chapter 3. Speeding Up Common Operations with ChatGPT

Everyone has to deal with some tedious, time-consuming, and repetitive tasks. One of these for me is making the bed every morning, not only mine (and my husband's) but also my children's. You must do it for the decorum of the house, but it's very dull, takes time, and is repetitive. To solve the problem, I could hire a housekeeper and ask them to arrange the beds for me. This is a bit like what ChatGPT does in the digital world. You can ask it to become your housekeeper and to carry out tedious and repetitive tasks for you. If there were a ChatGPT that could make beds, I would definitely buy it (of course, only at a reasonable cost).

Similarly, when you use Excel, you often have to deal with tedious, time-consuming, and repetitive tasks. You want to speed them up as much as possible. These include, for example, the creation of templates, filling tables with patterns defined as dates or progressive numbers, and even enriching a dataset with values external to the dataset. Well, ChatGPT can be your housekeeper.

In Chapter 2, you saw that ChatGPT can be a companion in your activities, an assistant, or a secretary—call it what you want. The idea is that ChatGPT assists you while you carry out your tasks. In this chapter, you will use ChatGPT in a new way: as a housekeeper, to carry out tedious, repetitive, and time-consuming tasks for you.

This chapter shows you how to use ChatGPT for:

- Creating templates
- Autofill
- Data enrichment

Let's start with creating templates.

Creating Templates

A *template* in Excel is a predefined file with a ready-made structure and format for specific tasks, such as budget management or inventory. Usually, a template contains

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

preconfigured tables, formulas, and graphs. When I refer to a template in Excel, I am thinking of the templates I see when creating a new workbook, such as budget planners and homework logs. You can make your own personal templates by creating an Excel workbook with the layout, formatting, and formulas you want, saving it as an Excel Template, and putting it in the appropriate folder. Once you've created the template, you can use it to create new workbooks based on the template, so you don't have to reinvent the wheel every time.

Using a template offers several advantages, such as saving time, achieving consistency in format and style between different documents, and reducing errors. In essence, a template can speed up repetitive tasks such as creating reports, invoices, or calendars by including predefined layouts, formatting, and formulas.

To create your template in ChatGPT, write a prompt with the story structure you learned about in Chapter 2: context, main point, and action. There are two application cases:

- Create the template from scratch, providing ChatGPT with all the details
- Create the template from a (handwritten) sketch, which you provide to ChatGPT as input

Let's analyze both cases, starting with creating the template from scratch.

Creating Templates from Scratch

When you create a template from scratch, you must provide ChatGPT with all the necessary information. Start a conversation with ChatGPT by following the flow described in Figure 3-1.

Figure 3-1. The flow to follow to generate an Excel template using ChatGPT

The prompt follows the story structure, where the final stage of the story is the creation of the Excel file, already included in the context prompt:

1. After logging in to *https://chatgpt.com*, start a new conversation, defining the context, which contains the fact that you want to create a template:

```
I want to create an Excel template. I'll provide you with all the options to
include in the template, and when I have finished, you'll format it as an
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
.xlsx file.
```

2. Specify the structure of the workbook, such as the number of worksheets, the number of tables, and the structure of the tables with columns. For example, you can create a single worksheet containing a single table using the following prompt:

```
Create a single worksheet containing a table with the following columns:
- Date
- Category
- Description
- Amount
- Payment Method
Format the table as an Excel table and name it MainTable.
Add 10 empty rows to the table.
```

   At this point, ChatGPT has generated a file in *.xlsx* format, which you can download. Next, open the file. Figure 3-2 shows the result.

Figure 3-2. The Excel file generated by ChatGPT after the second step

3. Define how to format the data, as shown in the following prompt:

```
Format cells Number as follows:
- Column A as a date (YYYY/MM/DD)
- Column B as text
- Column C as text
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
- Column D as currency with symbol $
- Column E as text
```

Download the file generated by ChatGPT. To test if the formatting is correct, enter the following information in the first row of the table:

```
2025-01-08, MyCategory, MyDescription, 3, Cash
```

The date refers to January 8, 2025. Excel automatically formats the date and amount, as Figure 3-3 shows. Once you have tested the formatting, you can delete the content of the first row.

Figure 3-3. The Excel file generated by ChatGPT after data validation

4. Add more details to the worksheet. For example, you can add a formula that automatically totals the Amount field, as shown in the following prompt:

```
Add a row below Amount and outside the table for total Expenses, calculated as
the sum of all the Amounts.
```

Figure 3-4 shows the Excel file generated by ChatGPT.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 3-4. The Excel file generated by ChatGPT after adding a formula to the worksheet

NOTE

This method calculates a total value using a cell formula. The goal of this example is to show that you can use ChatGPT to insert formulas outside of tables and enhance your worksheet with various formulas. Alternatively, you can use the Total Row feature, which is available for Excel tables.

Go to Table Design and check the box for Total Row. Excel will add a row at the bottom of the table with built-in functions. You can click on the cells in the total row to choose other functions from a drop-down menu, such as Sum, Average, Count, and more.

5. At this point, you can apply a style, defining colors, borders, and so on. For example, you can use the following prompt:

```
Apply the following style to the table MainTable:
First row:
- Set the text to bold
- Set the text color to black
- Set the background color to light gray
- Align text to the center

Other rows:
- Align text to the left for text columns and right for numeric columns
- Set the background color to white for all the cells starting from the
  second row
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
Borders:
- Apply thin black borders
```

Figure 3-5 shows the file generated by ChatGPT after applying the style.

Figure 3-5. The Excel file generated by ChatGPT after applying your style

WARNING

In some cases, ChatGPT may not generate the correct template. This is because the models behind ChatGPT are not perfect. To resolve this issue, try one of the following suggestions:

Rerun the exact same prompt several times until you get the desired result. For example, you can use the prompt: "Please try again."

Change the prompt slightly.

If you lose the cell's formatting, rerun the prompt, asking ChatGPT to apply specific formatting.

Exercise 1

Write prompts for ChatGPT to generate the table in Figure 3-6. Format the ISBN field as XXX-X-XXXXX-XXX-X. The calculation of Total Books should be the count of all the books in the table.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 3-6. The table to generate using ChatGPT

You can find the solution containing the prompt in the book's GitHub repository as *03/exercise1_solution.txt*.

So far, you have seen how to organize a conversation with ChatGPT step-by-step so that it generates the desired output. You've probably noticed how tiring it is to follow the conversation with ChatGPT. It would be great if you could give all the conversation prompts in one go and then let ChatGPT digest them gradually. Well, I have some good news for you. This is possible, which is exactly what we will see in the next section.

Automatically Generating a Template from Scratch

To automatically generate a template from scratch, write all the prompts in a single file and then feed the file to ChatGPT. You will need to define an opening prompt accompanying the file loaded into ChatGPT. The opening prompt specifies exactly how ChatGPT should behave when executing the instructions contained in the file.

For example, you can use the following opening prompt to define the task:

```
Open the attached file and identify the first instruction, which begins with the
keyword [STEP].

Carefully read the first instruction and follow the steps exactly as described.
Complete the first instruction fully before moving on to the next.
After completing the first instruction, move on to the next one and repeat the
process, following the sequence strictly.
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
Ensure that each instruction is executed separately without combining or
skipping steps.
If you encounter an unclear instruction or one that cannot be executed, stop the
process.
Provide a detailed explanation of the issue.

Once all instructions are completed, ensure the following:
- Every request has been implemented as specified.
- All required formatting and styles are properly applied.

After completing all steps, confirm that the process has been successfully
executed.
Share the generated file for review.

Execute each instruction as a separate unit, without skipping or combining
multiple instructions.
Follow a clear and logical sequence, ensuring each step is completed before
moving on to the next.
```

Attach to the opening prompt the *template_instructions.txt* file formatted as follows:

```
[STEP] Prompt 1

[STEP] Prompt 2

[STEP] Prompt N
```

You can test this technique directly using the file *03/template_instructions.txt*, found in the book's GitHub repository. This file contains the instructions described in the "Creating Templates from Scratch" section, preceded by the keyword `[STEP]`, as shown in the following prompt:

```
[STEP] I want to create an Excel template. I'll provide you with all the options
to include in the template, and when I have finished, you'll format it as an
.xlsx file.

[STEP] Create a single worksheet containing a table with the following columns:
- Date
- Category
- Description
- Amount
- Payment Method
Format the columns as an Excel table and assign the name "MainTable". Add 10
empty rows to the table.

[STEP] ...
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Using this technique can be time-consuming for ChatGPT, so don't worry if ChatGPT doesn't provide the answer immediately.

TIP

When using this technique, I've noticed that in most cases, ChatGPT fails to correctly format the table as an Excel table (despite it being explicitly asked to do so in the instructions). If this happens, manually pass it the following prompt after it has executed all the prompts contained in the template:

```
Format the generated table as an Excel Table.
```

After this statement, ChatGPT formats the table correctly.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

CHATGPT AS A FRIEND

I have tested this automatic technique many times with different templates and almost always obtained the same result: the tables were not formatted as Excel tables. Only after one or two trials did ChatGPT format the file correctly. I racked my brain, trying to understand why this happened. I changed the order of the prompts and the prompt text, but the result was still the same. Until I had a flash of genius, or as Gary Klein calls it in one of his books, *creative desperation*: I asked ChatGPT why it didn't consider the instruction I had given it. The response was shocking, beyond all expectations.

I gave it this prompt as input:

```
Explain why you didn't format the table as an Excel table and name it
MainTable
```

ChatGPT gave me the following response:

```
I did not apply the instruction to format the table as an Excel table and
name it MainTable for the following technical reasons:
  1. Difference between cell ranges and Excel tables: Although the data
     has been formatted as a range of cells with proper borders and styles,
     the specific Excel table format which includes features such as
     automatic filtering, dynamic rows, and the table name.
  2. Library used (openpyxl): In Python, the openpyxl library supports
     the creation of formatted tables (also called "Tables" in Excel),
     but the process requires specific configuration which has not been
     done. I should have used the openpyxl.worksheet.table.Table method
     to convert the range to a table and name it MainTable.
If you want, I can immediately correct this and update the file with the
correctly formatted table and the requested name. Let me know!
```

Aside from the first answer, which might seem bizarre, the second answer explains the technique ChatGPT uses to generate Excel files. In practice, ChatGPT transforms the prompts into Python code (Python is a programming language) and executes the code using the openpyxl library, which is precisely the Python library for generating Excel files. The performance and the quality of results depend on the chosen model. The Python library used for generating Excel files may also depend on the model. Some models use the XlsxWriter Python library, and others use the openpyxl library.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Creating Templates from Sketches

At the end of Chapter 2, I told you about the photo with the trees and the use of ChatGPT to extract a table from that photo. I also promised you that in the next chapter (this one), you would see how to create Excel tables directly from photos. Well, now it's time to talk about it.

Imagine that you have made notes on a piece of paper relating to the expenses you have incurred recently, as shown in Figure 3-7. (The figure is deliberately faded and not optimal for reasons I will explain shortly.)

Figure 3-7. A photo of a handwritten table

Now, imagine that you want to transcribe your paper table (and the total amount row) into an Excel sheet. The traditional approach involves manual transcription from paper to computer. But by using ChatGPT, you can speed up the operation.

You can proceed in two ways. In the first method, proceed using the ChatGPT web interface and follow these steps:

1. Before connecting to the ChatGPT web interface, take a piece of paper and write a table by hand. For example, you can copy the table in Figure 3-7 (typed out in Table 3-1). Also, add the Total Amount row, as shown at the bottom of Figure 3-7.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 3-1. An example of a table

Date	Category	Description	Amount	Payme Metho
2024/12/12	Grocery	Weekly Grocery	$43.00	Cash
2024/12/12	Presents	Present for Giulia	$12.00	Credit

2. Take a photo of your table and the total amount, as shown in the original photo in Figure 3-7. It doesn't matter whether the photo is excellent quality as long as it is legible, like the one in Figure 3-7. Don't worry about lights, shadows, or taking perfect photos.
3. Upload the photo to ChatGPT via the upload button and use the following prompt:

```
Format the attached table as an Excel file and make it downloadable.
Follow this workflow to build the Excel file:
Populate columns and rows
Format as an Excel Table except for the row Total Amount
Calculate the Total Amount using a formula (Sum of the Amount column)
```

4. Download the Excel file produced by ChatGPT. You should get something very similar to Figure 3-8.

Figure 3-8. The Excel file produced by ChatGPT

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

5. As a bonus, you can also ask ChatGPT to add more consistent lines. This is particularly useful if you want to generate example files for other cases. Use this prompt, continuing the conversation with ChatGPT:

```
Populate this table with 15 consistent rows.
```

Figure 3-9 shows a possible result generated by ChatGPT.

Figure 3-9. The result of this prompt: Populate this table with 15 consistent rows

If you have followed all the steps described previously, you may have noticed a minor complication. You probably took the photo with your smartphone, uploaded it to the cloud, downloaded it on your computer, and finally uploaded it to ChatGPT. This operation is cumbersome and slow.

There is an alternative way to proceed: the ChatGPT app for smartphones! You can download the official ChatGPT app from your favorite app store, log in to your account, and have ChatGPT accessible from your smartphone. Everything I have said so far about the web interface also applies to the smartphone app. So you can decide from time to time whether to use the web interface or the app.

In this specific case, I suggest you use the app when creating the template from a photo. After logging in to the app, you have an interface available where you can directly take the picture, as shown in Figure 3-10.

After taking the photo directly in ChatGPT, perform steps 3 and 4 described previously. If you don’t want to download the Excel file to your smartphone, you can continue the

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

conversation via the web interface.

TIP

An interesting aspect of ChatGPT is that it can store conversations, which you can access via a sidebar on the left. These conversations are available in both the app and the web version, so you can start a conversation in the app and continue it in the web interface and vice versa. This option is not available if you start a temporary chat.

Figure 3-10. The ChatGPT smartphone app interface with the ability to upload photos

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Exercise 2

If you haven't recycled it yet, find the latest supermarket flyer you received in the mail and select a page of your choice. Take a photo and input it to ChatGPT using the following prompt:

```
Format the attached table as an Excel file and make it downloadable.
Follow this workflow to build the Excel file:
Populate columns and rows with the items in the flyer
Format as an Excel Table
```

I tried this exercise, and the result was quite good. ChatGPT organized the items into a table with the structure shown in Table 3-2 and identified almost all items on the page.

Table 3-2. The table structure generated by ChatGPT from the flyer page

Category	Product Name	Size/Quantity	Price (€)

From the experience of this exercise, I always advise you to check the accuracy of the results produced by ChatGPT.

At this point, you should have a pretty clear idea of how to use ChatGPT to generate templates, either from scratch or from a photo or image. In the next section, you will learn about the concept of Autofill and how to use ChatGPT to speed up this operation.

Autofill

Autofill in Excel is a function that allows you to automatically fill a series of cells with values, formulas, or patterns based on a predefined template. You can use Autofill to complete number sequences, days of the week, months, formulas, and more.

Suppose you have the table shown in Table 3-3 (available in the book's GitHub repository as *03/autofill.xlsx*).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 3-3. An example table

#	Date	Day of Week
1	2024-01-01	
2	2024-01-02	
3	2024-01-03	

The first column is a progressive number, the second is a progressive date in the format YYYY-MM-DD, and the third is the day of the week. It is currently empty, but you would like it filled with the day of the week corresponding to the date. In the example, the first date should correspond to Monday, the second to Tuesday, and so on, given that January 1, 2024, was a Monday.

Suppose you want to extend the table using Autofill to have dates up to January 8, 2024. Let's see how to perform this operation first manually using only Excel and then with ChatGPT.

To proceed manually, do the following:

1. Select the first three dates and move the cursor to the bottom right, where there is a small square called the Fill Handle, as shown in Figure 3-11.
2. Move the mouse cursor until it becomes a black cross, then click and drag down a few rows until the date 2024-01-08 appears, as shown in Figure 3-12.
3. Release the cursor. All dates are populated automatically through 2024-01-08, as shown in Table 3-4.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 3-11. How to activate Autofill in Excel (step 1)

Figure 3-12. How to activate Autofill in Excel (step 2).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 3-4. The result of the Autofill process

#	Date	Day of Week
1	2024-01-01	
2	2024-01-02	
3	2024-01-03	
	2024-01-04	
	2024-01-05	
	2024-01-06	
	2024-01-07	
	2024-01-08	

Exercise 3

Repeat the same operations just described for the `#` column containing the progressive number, and the `Day of Week` column. To use the Autofill function in the latter case, remember to write the values for at least the first two rows.

At this point, you are ready to repeat the same operations using ChatGPT. Remember that each prompt must be structured into three parts: context, main point, and action. In our example, we directly load the *autofill.xlsx* dataset into ChatGPT, and this forms our context. Then we ask ChatGPT to fill in the missing values (main points). Basically we use ChatGPT to perform the Autofill for us automatically. Finally, we tell ChatGPT to format the output as an Excel table, maintaining the style of the original file.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

TIP

Using ChatGPT to perform Autofill is particularly useful for populating a dataset with many values. If you want to fill in only a few values, you can still use ChatGPT, but the advantage over the manual solution is minimal.

Use the following prompt:

```
Consider the table attached.

Using Autofill, populate all the table columns with dates starting from
2024-01-01 to 2024-01-08. Use the correct day of week (Monday, ...).

Format the resulting table as an Excel table and keep the original table layout.
```

ChatGPT performs the operations and then generates the file in Excel format for you to download. Figure 3-13 shows a portion of the file generated by ChatGPT.

Figure 3-13. The file generated by ChatGPT after running the Autofill

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

In some cases, the file generated by ChatGPT may not contain data formatted as a table and with the same layout as the original file. In this case, continue the conversation with ChatGPT, repeating the last part of the prompt:

```
Format the resulting table as an Excel table and keep the original table layout.
```

ChatGPT should generate the correctly formatted file after a couple of tries, at most.

As you've seen, you can use Autofill to automatically fill the rows of a table with predefined values or sequences. If you have a lot of values to fill in, use ChatGPT to speed up this operation. But what if you want to extend your data with data external to your source dataset? You can still use ChatGPT to do this. Let's see how in the next section.

Data Enrichment

Data enrichment consists of extending an original dataset with data from other sources. You can extend a dataset with geographic data, descriptions, or other information from external sources, such as the web, or internal sources, such as other datasets from the same source. To extend a dataset with other data, the dataset must not contain duplicates and must have at least one column to be used as a unique identifier.

Let's take an example. Suppose you have a workbook with two tables organized into two separate worksheets. You can download the workbook from the book's GitHub repository as *03/data_enrichment.xlsx*. The first table, called `Books` (available in the Books tab), contains the book title and status (whether it is currently being read, is set to read, or has already been read), as shown in Table 3-5.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 3-5. The `Books` table in the data_enrichment.xlsx example file

Title	Status
Effective Data Storytelling	Read
Become a Great Data Storyteller	To read
Storytelling with Data	To read
The Data Storyteller's Handbook	Reading

The second table, called `Info` (available in the Info tab), contains the book title, author, and year of publication, as shown in Table 3-6.

Table 3-6. The `Info` table in the data_enrichment.xlsx example file

Title	Author	Year
The Data Storyteller's Handbook	Kat Greenbrook	2023
Storytelling with Data	Cole Nussbaumer Knaflic	2015
Effective Data Storytelling	Brent Dykes	2019
Become a Great Data Storyteller	Angelica Lo Duca	2025

Suppose you want to extend the first table with the information contained in the second. The two tables share a common column, `Title`, which will act as a unique identifier for matching between the two tables.

Use the function `VLOOKUP()` to extend the first table with the data from the second. The function `VLOOKUP()` has the following syntax:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
=VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])
```

`lookup_value`

The value you want to look for in the first column of the `table_array` table.

`table_array`

The table or range of cells to search for the value. The first column of this range is where `VLOOKUP()` searches for the `lookup_value`.

`col_index_num`

The number of the column (in the table range) from which to take the value to be returned. The first column always has the number 1.

`range_lookup` *(optional)*

Specify the match type. If set to `TRUE` (or omitted), it looks for an approximate match (values should be sorted in ascending order). If set to `FALSE`, it searches for an exact match.

The function returns the value in the column specified by `col_index_num`. Apply the function `VLOOKUP()` to the `Books` table as follows:

1. Select cell C2 in the Books worksheet, as shown in Figure 3-14.

Figure 3-14. How to enrich data using `VLOOKUP()` (step 1)

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

2. Write the following formula to select the author:

```
=VLOOKUP([@Title],Info[#All],2,FALSE)
```

3. Then press the Enter key. Figure 3-15 shows the result.

Figure 3-15. How to enrich data using `VLOOKUP()` (step 2)

4. Change the name of the column to `Author`.
5. Repeat the operation for the `Year` column (using a formula). Figure 3-16 shows the final result. You can find the solution in the book's GitHub repository as *03/data_enrichment_manual.xlsx*.

Figure 3-16. The final result after data enrichment

So far, you have seen how to extend the dataset manually. This example required that I already have the enriched data in a separate table. However, you may not have the data readily available and you want to extract it from an external source, such as the web. In this case, ChatGPT can help you.

To demonstrate how to extract data from the web, we'll use a simplified version of the dataset, which contains only the `Books` table, and we'll ask ChatGPT to fill it in using the data it knows. You can find this simplified version in the book's GitHub repository as *03/data_enrichment_simplified.xlsx*.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

First, let's define a prompt that consists of three parts: context, main point, and action. As context, we use the *data_enrichment_simplified.xlsx* dataset, which we load into ChatGPT using the appropriate button to load data. Let's start the conversation with ChatGPT by writing the following prompt:

```
Consider the table Books in the first worksheet of the attached file. Extend the
table Books with the following columns: author, year of publication, publisher.
Extract the required information from the web. Format the resulting table as an
Excel Table. Keep the original table layout and style.
```

ChatGPT generates the result shown in Figure 3-17, but it is not formatted as an Excel file. Furthermore, the table does not contain some data because it is not available in the GPT-4o model used by ChatGPT in this experiment.

Figure 3-17. The table enriched by ChatGPT does not contain all the required information

NOTE

In some cases, ChatGPT may not allow you to download the generated file because it remains temporarily stored in memory. If you attempt to download the file after a long period of inactivity, it may have already been deleted and is no longer available.

In other cases, ChatGPT might encounter issues during the file generation or download process, such as a timeout, server error, or browser-related restriction. To avoid losing access to the file, download it as soon as it's created.

To solve the problem, ask ChatGPT to repeat the operation by searching for information on the web. Select the "Web search" button in the ChatGPT interface, as shown in Figure 3-18.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 3-18. The Search button in the ChatGPT web interface

Once the button is enabled, write this prompt:

```
Repeat the operation.
```

ChatGPT now generates a new table, as shown in Figure 3-19.

Figure 3-19. After ChatGPT uses the Search function, the enriched table contains all the required information

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

The table is not yet downloadable as an Excel file. So, ask ChatGPT to format the table as an Excel file, keeping the original formatting and data present in the source file:

```
Format the resulting table as an Excel Table. Keep the original table layout and
style.
```

WARNING

In some cases, ChatGPT can *hallucinate*, meaning it generates incorrect results. Always remember to check the results produced by ChatGPT before using them!

At this point, the file is ready, as shown in Figure 3-20.

Figure 3-20. The final file generated by ChatGPT

Compared to the original file, ChatGPT did not maintain the original theme of the table but still enriched the data correctly.

From this example, we learned the following about ChatGPT:

- It automates most operations, such as calculations and data enrichment.
- Sometimes you need to ask it several times to format the table as an Excel table and keep the original style.
- It doesn't always maintain the original table's theme and style.

TIP

ChatGPT sometimes fails to generate the file to download and gives instructions for you to create the file yourself. In this case, you can ask ChatGPT to retry, for example, by using the following prompt: "Please try again." Alternatively, you can follow the instructions provided by ChatGPT and apply them by hand to your original Excel file.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Exercise

Write instructions for ChatGPT to populate Table 3-7 with the country name and geographic coordinates.

Table 3-7. The table to enrich

City	Country	atitude	ongitude
Pisa			
New York			
Sydney			
Paris			

The Excel file containing the table is available in the book's GitHub repository as *03/exercise4.xlsx*, and the solution is here: *03/exercise4_solution.txt*.

NOTE

I conclude this chapter with a personal reflection. Last night, I had a moment of pure contemplation: I stopped to observe my nine-year-old daughter Giulia while she played the piano. She was enchantingly graceful. The melody that filled the room was so sweet and enveloping that it touched the deepest strings of the soul. At that moment, a thought took shape in my mind. ChatGPT, upon closer inspection, is like a piano: it has infinite keys and can create extraordinary melodies. However, just as a piano remains silent and inert without someone who can play it, ChatGPT remains unused if you don't approach it with the proper prompts. It is human interaction, awareness, and creativity that give it voice and make it extraordinary.

Summary

In this chapter, you learned how to speed up some common operations using ChatGPT:

A template is a predefined file with a ready-made structure and format for specific tasks. You can use ChatGPT to speed up template creation. Structure the prompt following a precise order: first, define the context and then the

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

design of the workbook, with the number of tables and their structure, data validation, additional elements, and style.

With ChatGPT, you can create a template from scratch or use a photo of a handmade sketch. You can even ask ChatGPT to extract Excel tables from flyers or other images.

Autofill in Excel is a useful function for automatically filling in a series of cells with values, formulas, or patterns based on a predefined template. ChatGPT can speed up the autofilling process if the amount of data to insert is large.

Enriching a dataset involves extending an original dataset with data from other sources. Use ChatGPT's Search function to enrich a dataset with information extracted from the web.

In the next chapter, you will learn how to use ChatGPT to perform advanced operations.

Activity

Solve the word puzzle in Figure 3-21 by circling all the words you see. In the end, you will discover something about Autofill by reading the letters that are not circled, which will now form words. You can find the solution in Appendix D.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 3-21. Word puzzle to solve

Words to search for:

AUTOFILL

CHATGPT

CONTEXT

ACTION

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

EXCEL

MAIN POINT

PROMPT

ROW

TABLE

TEMPLATE

WEB SEARCH

Word puzzle solution. Autofill:

Chapter 4. Advanced Operations with ChatGPT

When I started programming in the early 2000s, fixing a bug in the code was a real challenge. If something wasn't working, you didn't even know where to start. It could take hours, sometimes days, of trying to check line by line in the hope of finding that bug in the code. Then, suddenly, a colleague or friend would arrive and, with a quick glance, identify the problem and point it out to you as if it were the most obvious thing in the world. And you, caught between disbelief and gratitude, would thank them heartily.

Then came the internet, Web 2.0, and programming forums. If your code threw errors, all you had to do was use a search engine, and sooner or later, you would find someone who had had the same problem and had already solved it. Debugging times have shortened dramatically.

Today, with ChatGPT, everything is even faster. Just paste the code and ask where the error is. You'll have the solution in a few minutes, sometimes in a few seconds. When I think back to 20 years ago, when I could have spent entire days on a single bug, the progress is impressive. But it doesn't end there. With generative AI, you no longer have to write code; ChatGPT does it directly. Just describe the problem, and it returns the code ready to use, both for programming in languages such as Python, Java, or C# and for creating advanced and efficient formulas in Excel. And that's exactly what you'll see in this chapter. More precisely, this chapter covers the following topics:

- Using ChatGPT to analyze a dataset
- Using ChatGPT for PivotTables
- Combining ChatGPT and macros
- Using ChatGPT to write VBA code

Let's start with the first one, using ChatGPT to analyze a dataset.

Using ChatGPT to Analyze a Dataset

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

The previous chapter showed how to use AI to help you write or understand formulas. However, having formulas is not enough, and knowing how to use them is not enough. You also need to use formulas to perform the proper data analysis. For example, you might have a beautiful dataset and know all the Excel formulas, but you have no idea what to ask of that dataset and what formula to use to extract something important from your data. Well, AI and, in particular, ChatGPT can help you with this. In this section, you'll see how AI can analyze the dataset and generate advanced formulas. To analyze the dataset, follow the flow described in Figure 4-1.

Figure 4-1. The workflow for analyzing a dataset

A typical workflow in more detail might go like this:

1. Start with your dataset in Excel. Extract a portion containing the column names and possibly some example rows (or use the entire dataset if it is not too large).
2. Upload it to ChatGPT. (Alternatively, as seen in previous chapters, you could give ChatGPT context about the dataset.)
3. Ask ChatGPT to extract the category from your dataset. Examples of categories include finance, environment, health, society, and so on. This step is essential to help ChatGPT take the next step.
4. Ask ChatGPT to define relevant questions for your dataset. For example, Table 4-1 shows some possible questions to ask about the data based on the category to which it belongs.
5. Once you've identified the questions, ask ChatGPT to answer them using Excel formulas.
6. Tell ChatGPT to implement the responses into your original dataset and make it downloadable.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 4-1. A list of possible questions based on the data category

Category	Requests
Health	What are the main risk factors for a particular disease?
	How do diagnoses vary over time and space?
	How effective is a specific treatment?
Finance	What are the trends in the financial markets?
	What factors influence stock prices?
	What is the average return on an investment portfolio?
Environment	What are the primary sources of pollution in a region?
	How does air quality change over time?
	What measures can reduce the environmental impact?
Social media	What are the trends in online conversations?
	How does user engagement vary across different platforms?
	What is the general sentiment on a particular topic?
Education	What are the factors that influence student achievement?
	How does the school dropout rate vary?
	What teaching strategies improve learning?
E-commerce	Which products are best-selling in a given period?
	What factors influence purchasing behavior?
	Which marketing strategies are most effective?

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Category	Requests
Transport	What are the most common traffic patterns in a city?
	What are the main factors causing transport delays?
	How do we optimize routes to reduce fuel consumption?
Energy	What are the most efficient energy sources?
	How does energy consumption vary over time?
	What strategies can reduce energy waste?
Manufacturing	What are the most common causes of failures in production lines?
	How can we improve assembly line efficiency?
	Which materials offer the best value for money?
Sport	What factors influence athlete performance?
	Which training strategies are most effective?
	How can we predict the risk of accidents?'

Let's apply this workflow to a full example.

Dataset

Consider the synthetic dataset available in the book's GitHub repository at *04/sales_dataset.xlsx* and shown in Figure 4-2.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 4-2. An extract from the sales dataset

The dataset shows product sales and contains the following fields:

`Transaction ID`

: Unique identifier for each sale

`Date of Sale`

: Date of the transaction (YYYY-MM-DD)

`Product ID`

: Unique identifier for the product sold

`Product Name`

: Name or description of the product

`Quantity Sold`

: Number of units sold

`Unit Price`

: Selling price per unit

`Total Sale Amount`

: Total revenue from the sale (Quantity × Unit Price – Discounts)

`Payment Method`

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Cash, credit card, PayPal, etc.

`Store Location`

City, region, or country where the sale occurred

`Customer Type`

New, Returning, VIP, etc.

Extract the Category and Relevant Questions

Start a new conversation with ChatGPT, load the dataset, and execute the following prompt:

```
Consider the attached dataset describing an Excel table named MainTable. Identify
the category to which the dataset belongs (finance, health, and so on) and list
the top 5 helpful questions to analyze this dataset.
```

In our example, ChatGPT associated the dataset with the Sales/Retail/Business category and generated the following questions:

`Which product generates the highest revenue?`

`What are the trends in sales over time?`

`Which store location has the highest total sales?`

`What is the distribution of payment methods used by customers?`

`How does customer type (New, Returning, VIP) impact sales and revenue?`

By repeating the operation several times and starting a new conversation each time, you'll notice that the questions change slightly, but overall, they are generally the same.

TIP

You can ask ChatGPT to generate five more questions by writing the following prompt:

```
List 5 more questions.
```

Repeat to get five more questions each time.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Generate Answers

Now ask ChatGPT to generate the answers to the questions using Excel. Use the following prompt:

```
Write the Excel formula or steps to answer each question and how to implement it
in Excel. Don't use PivotTables. Use the table name MainTable and column names
to write formulas. Write the general formula or steps to follow whenever
possible without focusing on a specific value.
```

You specifically ask ChatGPT not to use PivotTables because, as we will see later in this chapter, in its current version (as of this writing in early 2026), ChatGPT cannot apply PivotTables to an Excel file independently. However, if at the time you are reading this, you use a version of ChatGPT that is capable of producing PivotTables, feel free to delete the text "Don't use PivotTables" from the prompt.

For each answer, ChatGPT lists the steps to implement it in Excel. For example, to find the answer to Question 1, "Which product generates the highest revenue?" ChatGPT generated the following response:

```
To calculate total revenue for each product, use SUMIFS:

=SUMIFS(MainTable[Total Sale Amount], MainTable[Product Name], A2)


Replace A2 with the product name cell reference.

Copy the formula for all unique product names.

Steps:

List unique product names in a new column.

Use the SUMIFS formula to sum Total Sale Amount for each product.

Identify the highest value using MAX or by sorting the revenue column
in descending order.
```

First, ChatGPT proposes the formula to calculate the answer and then provides the steps to implement it. In the example, ChatGPT proposes the use of the function `SUMIFS`, which sums the values in a range that satisfy multiple criteria. The syntax is as follows:

```
SUMIFS(sum_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Let's unpack that:

`sum_range`: The range of values to sum

`criteria_range1`: The first range to apply a condition to

`criteria1`: The condition to be met in `criteria_range1`

`criteria_range2, criteria2 (optional)`: Additional conditions to filter the sum

In the example generated by ChatGPT, the function `SUMIFS` sums the values of the `Total Sale Amount` column of the `MainTable`, but only for the rows in which the value of the `Product Name` column corresponds to the value present in cell A2. ChatGPT, therefore, does not directly answer the question "Which product generates the highest revenue?" but selects the relevant data (and processes it) to help you answer the question more easily.

Implement Answers

At this point, ask ChatGPT to implement the answers it gave in the original file. Use the following prompt:

```
Add a sheet to the original file for Question 1 named as the question. Format the
original dataset as an Excel Table and name it MainTable. Implement all required
steps to answer Question 1. If the response involves a formula, apply it in the
corresponding Excel sheet. Ensure that calculated values are replaced with
formulas and all the steps you have listed to answer the question are
implemented. Make the file downloadable.
```

In this case, ChatGPT generated an Excel file, which you can download from the book's GitHub repository at *04/sales_dataset_q1.xlsx*. Figure 4-3 shows the portion of the generated file related to Question 1. Interestingly, the values in the `Total Revenue` column are calculated as formulas.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 4-3. The worksheet generated by ChatGPT in response to Question 1

To answer the question “Which product generates the highest revenue?” you need to do a little extra work. Proceed as follows:

1. With the mouse, select all the non-empty cells of the Top Revenue Product worksheet, and from the Home tab of the Ribbon, select Format as Table, choose a format, and confirm the range and headings in a dialog box (Figure 4-4).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 4-4. Select the non-empty cells of the worksheet that answers Question 1 and format them as a table

2. Select the arrow at the right of the `Total Revenue` column and click Sort Largest to Smallest (Figure 4-5). Then click OK, if needed.

Figure 4-5. Select Sort Largest to Smallest

3. The first product in the table answers the question "Which product generates the highest revenue?" In our case, the product that produces the most revenue is Headphones, as shown in Figure 4-6.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 4-6. The first row of the table shows the product that generates the highest revenue

After using the procedure shown to answer the first question, we can make some conclusions:

Use ChatGPT to build the framework for answering your data questions.

Use your knowledge and experience to answer the questions based on what ChatGPT generated.

ChatGPT, therefore, automates basic and repetitive operations, but the final answer is always up to you.

After you answer the first question, ask ChatGPT to answer the remaining questions, one at a time. Use the following prompt to answer the second question:

```
Repeat for Question 2. Use the Excel file you have generated in the previous
step, ensuring that all the formulas previously calculated are maintained.
```

Exercise 1

Check if the answer ChatGPT provided to Question 2, "What are the trends in sales over time?" is correct. If necessary, refine your answer. You can find the answer to Question 2 in the book's GitHub repository at *04/sales_dataset_q2.xlsx*. As in the answer to Question 1, in this case ChatGPT generated the framework to answer the question but has not given a definitive answer. You can find the answer by, for example, inserting a graph showing the sales trend.

Follow these steps:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

1. Turn non-empty cells into a table (as described in the preceding section "Answers Implementation").
2. Insert a graph visually showing the trend: select the Insert tab from the Ribbon and then select Insert Line or Area Chart → 2-D Line → Line (Figure 4-7).

Figure 4-7. Insert a graph to answer the question "What are the trends in sales over time?"

Challenge: Ask ChatGPT to generate the graph and insert it into the output file.

Extending to All Questions

After answering Question 2, repeat the process for Questions 3, 4, and 5. For example, to address Question 3, enter this prompt:

```
Repeat for Question 3. Use the Excel file you have generated in the previous
step, ensuring that all the formulas previously calculated are maintained.
```

ChatGPT adds a worksheet to the file whenever it answers a question. You can download the files generated for this example from the book's GitHub repository at *04/sales_dataset_q3.xlsx*, *04/sales_dataset_q4.xlsx*, and *04/sales_dataset_q5.xlsx*. *qx* indicates the question number. The file relating to question *qx* contains all answers from 1 to *x*. Figure 4-8 shows the structure of the final file (*sales_dataset_q5.xlsx*), which answers all five questions.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 4-2 shows the questions and the process ChatGPT followed to answer them in this example.

Figure 4-8. ChatGPT answered each question using a separate worksheet.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 4-2. Questions and answers relating to the sales dataset

Question	Answer	Formula used
Which product generates the highest revenue?	List unique product names and use `SUMIFS` to calculate the total revenue for each.	`=SUMIFS(MainTable[Total Sale Amount], MainTable[Product Name], A2)`
What are the trends in sales over time?	List unique dates and use `SUMIFS` to calculate total sales per date.	`=SUMIFS(MainTable[Total Sale Amount], MainTable[Date of Sale], A2)`
Which store location has the highest total sales?	List unique store locations and use `SUMIFS` to calculate total sales per store.	`=SUMIFS(MainTable[Total Sale Amount], MainTable[Store Location], A2)`
What is the distribution of payment methods used by customers?	List unique payment methods and use `COUNTIF` to count occurrences.	`=COUNTIF(MainTable[Payment Method], A2)`
How does customer type (New, Returning, VIP) impact sales and revenue?	List unique customer types and use `SUMIFS` to calculate total revenue per type.	`=SUMIFS(MainTable[Total Sale Amount], MainTable[Customer Type], A2)`

In most cases, ChatGPT proposed using the `SUMIFS()` function. You could have achieved a similar result using PivotTables. However, there are differences between `SUMIFS()` and PivotTables. Table 4-3 shows some pros and cons of `SUMIFS()` versus a PivotTable.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 4-3. A quick comparison between `SUMIFS()` and a PivotTable

	Pros	Cons
`SUMIFS()`	Formula automatically updates without having to update the table manually.	The formula `SUMIFS()` can become complex and difficult to manage in sheets with many conditions.
PivotTable	It allows you to analyze and summarize large amounts of data quickly. It can calculate sums, averages, counts, and other operations without writing complex formulas.	It is not dynamic in real time: if the data changes, it must be updated manually.

Both are useful and allow you to achieve similar results. It is up to you, based on the pros and cons, to decide which one to use.

At the time of writing this chapter, ChatGPT can't add PivotTables automatically to an Excel file. However, you can still use ChatGPT as a teacher to learn how to build a specific PivotTable. Let's see how in the next section.

Using ChatGPT for PivotTables

A PivotTable in Excel is a tool that helps you quickly summarize and analyze data. You can use PivotTables to easily organize, group, and calculate data (like sums or averages) by dragging and dropping fields, making it easier to find patterns and insights. Use ChatGPT as a teacher to describe the steps to implement a PivotTable.

Consider the following example to understand how to proceed. We will find the answer to Question 1 "Which product generates the highest revenue?" using PivotTables. Continuing the conversation in the same chat from which you downloaded the various Excel files with the answers, ask ChatGPT for the steps to implement Question 1 using PivotTables. Use the following prompt:

```
Describe the steps to implement Question 1 using PivotTables.
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

In my case, ChatGPT provided the following response:

```
Step 1: Select Your Data
  1. Open your Excel file containing the dataset.
  2. Click anywhere inside the dataset (MainTable) to automatically detect data
     for the PivotTable.
Step 2: Insert a PivotTable
  1. Go to the Insert tab on the Excel ribbon.
  2. Click PivotTable.
  3. In the Create PivotTable window:
     • Ensure the Table/Range is set to MainTable (if already formatted as a
       table).
     • Choose New Worksheet for the PivotTable location.
     • Click OK.
Step 3: Configure the PivotTable Fields
  1. In the PivotTable Fields panel, drag and drop:
     • "Product Name" into the Rows area.
     • "Total Sale Amount" into the Values area.
  2. Ensure that the "Total Sale Amount" field is set to Sum
     (Right-click → Summarize Values By → Sum).
Step 4: Sort to Identify the Top Product
  1. Click any value in the Total Sale Amount column inside the PivotTable.
  2. Go to the Data tab and click Sort.
  3. Choose to Sort by Total Sale Amount in Descending Order
     (Largest to Smallest).
  4. The top product will now be the first entry in the table.
Step 5: Apply Formatting (Optional)
  1. Click on the Total Sale Amount column.
  2. Go to the Home tab → Click Number Format → Select Currency or Number
     formatting.
  3. Adjust column widths for better readability.
Step 6: Identify the Product with the Highest Revenue
  • The product with the highest total revenue will be the first row in the
    sorted PivotTable.
  • Optionally, use Conditional Formatting (Home → Conditional Formatting → Data
    Bars or Color Scales) to highlight the highest revenue product.
```

By following the steps described by ChatGPT, you get the PivotTable shown in Figure 4-9, which is very similar to the table in Figure 4-6, although the formatting is much better (thanks to steps 5 and 6).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 4-9. The PivotTable created using the steps suggested by ChatGPT

In providing the answer, ChatGPT also added the following bonus:

```
Bonus: Create a Chart
  1. Click anywhere inside the PivotTable.
  2. Go to the Insert tab → Select Column Chart or Bar Chart.
  3. The chart will visually display the revenue distribution by product.
```

With this bonus, ChatGPT explains how to add a chart. By following the steps described, you obtain the graph shown in Figure 4-10.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 4-10. The column chart relating to the PivotTable for Question 1

So far, you have seen how to use ChatGPT to analyze a dataset and, as a teacher, to help build PivotTables. However, ChatGPT can go further by generating macros and VBA code. Let's see how in the next section.

Macros and VBA

A *macro* in Excel is a set of automated commands that allows you to quickly perform repetitive tasks, such as formatting data, creating reports, or performing complex calculations. You can write macros using the VBA (Visual Basic for Applications) programming language or record macros without writing code.

Before using macros, you must activate them following these steps (in Windows), if they are not already activated:

1. Open Excel and go to File → Options.
2. Select Trust Center in the left menu.
3. Click Trust Center Settings.
4. In the window that opens, select Macro Settings.
5. Choose the "Disable VBA macros with notification" option (recommended).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

6. Click OK to confirm your changes.

If you use macOS, follow these steps: Open Excel and go to Excel → Preferences → Security → Disable all macros with notification.

At this point, customize the Ribbon to display the Macros commands. Follow these steps to enable the Developer tab, which also contains macros:

1. Choose File → Options (Windows) or Excel → Preferences (macOS).
2. Select Customize Ribbon (Windows) or Ribbon & Toolbar (macOS)
3. Under Customize the Ribbon, check the Developer option and click OK (Figure 4-11).
4. Select Customize Ribbon → Main Tabs → Developer to add the Developer tab to the Ribbon.

Figure 4-11. The pane to add the Developer tab to the Ribbon

Once you have performed these operations, you will find the Developer tab in the Ribbon, which should appear as shown in Figure 4-12 (in Windows) or Figure 4-13 (in

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

macOS).

Figure 4-12. The Developer tab in the Ribbon (Windows)

Figure 4-13. The Developer tab in the Ribbon (macOS)

At this point, you are ready to use macros. We will look at two strategies. The first is to record a macro, and the second is to write the VBA code to generate the macro. Let's start with the first step: recording a macro.

Recording a Macro

Let's use an example to understand how macros work. Open the file *04/sales_dataset_q5.xlsx*, which was the result of the previous exercise (Figure 4-8). Inside the workbook, in addition to the main worksheet containing the main dataset, you will find five worksheets, one for each question. Each worksheet includes ranges of values not formatted as Excel tables (Figure 4-6). Let's now write a macro to transform the data contained in each worksheet into a table. We will record the macro for a single worksheet without writing code in VBA and then apply it to all other worksheets.

To record an Excel macro that converts the second worksheet into a table and formats the revenue column as currency, follow these steps:

1. Navigate to the second worksheet, Top Revenue Product.
2. Go to the Developer tab and click Record Macro. In the Record Macro dialog box:
    a. Enter a Macro Name (e.g., ConvertToTable).
    b. Choose where to store the macro (This Workbook).
    c. Optionally, assign a shortcut key (e.g., Ctrl + Shift + T). Including the Shift key prevents you from overriding the Ctrl + T shortcut that converts the selected cells into a table.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

   d. Add a description (e.g., “Converts a cell range into a table and formats revenue as currency”).

   e. Click OK to start recording.

3. Click cell A2 on the second worksheet. Press Ctrl + T to open the Create Table dialog.

   a. Ensure the “My table has headers” checkbox is checked.

   b. Click OK to create the table.

4. Identify the column containing revenue (e.g., `Total Revenue`, column B) and manually select all the rows for that column in the table.

5. Go to the Home tab, and in the Number group, click the dropdown and select Currency. Adjust the currency symbol and decimal places using the Format Cells option if needed.

6. Go back to the Developer tab and click Stop Recording.

7. Save the workbook as a Macro-Enabled Workbook (*.xlsm* format) to retain the macro. You can find the file in the book’s GitHub repository under *04/sales_dataset_q5.xlsm*.

Figure 4-14 shows the result after recording the macro: the data is formatted as an Excel table, and the `Total Revenue` column is formatted as currency.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 4-14. The data is formatted as an Excel table after recording the macro

If you were to apply the macro as it is to the other worksheets, though, you would get the wrong result, as shown in Figure 4-15. As you can see, only the first six rows were converted into an Excel table, while the others remain unchanged. This is because the macro was recorded using a six-row table. You don't need to try this out yourself. It's just to show you what would happen.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 4-15. If you apply the macro recorded for worksheet 2 to the other worksheets, you get an incorrect result because the macro records the specific values of the worksheet in question

To make the macro work correctly for other worksheets, you must generalize it, removing specific references to the data in worksheet 2. To do this, you need to manually modify the macro code. ChatGPT helps us with this. You'll use this procedure: open the VBA editor to access the macro, copy the macro code, paste it into ChatGPT, and instruct it to generalize. Follow these steps:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

1. Go to the Developer tab, click Macros, select ConvertToTable, and click Edit to open the VBA Editor.
2. Once the VBA Editor is opened, you will see the macro code in the module window (Figure 4-16).

Figure 4-16. The VBA Editor in Excel

Going into detail, recording the macro generated VBA code similar to this:

```
Sub ConvertToTable()
'
' ConvertToTable Macro
' Converts a sheet into a table and formats revenue as currency
'
' Keyboard Shortcut: Ctrl+Shift+T
''
    Range("A2").Select
    Application.CutCopyMode = False
    ActiveSheet.ListObjects.Add(xlSrcRange, Range("$A$1:$B$6"), , xlYes).Name = _
        "Table2"
    Range("Table2[Total Revenue]").Select
    Selection.NumberFormat = "General"
    Selection.NumberFormat = "[$$-en-US]#,##0.00"
End Sub
```

The macro defines the subroutine `ConvertToTable` using the keyword `Sub`. In VBA, subroutines are delimited by the keywords `Sub` (to get started) and `End Sub` (to finish). In VBA, comments are identified by the apostrophe (`'`). Any text that follows an

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

apostrophe in the same line is ignored by the compiler and is useful for documenting the code and improving its readability.

After the comments header, the cell A2 is selected with this instruction:

```
Range("A2").Select
```

This selection indicates the data to use. The next instruction is this:

```
Application.CutCopyMode = False
```

It turns off the cut/copy mode, preventing any copy/paste operations from interfering with the macro. At this point, the macro creates a `ListObject` table using the following command:

```
ActiveSheet.ListObjects.Add(xlSrcRange, Range("$A$1:$B$6"), , xlYes) _
.Name = "Table2”
```

This line converts the range A1:B6 into a table, assigning it the name `Table2`. The parameter `xlYes` indicates that the first row of the range contains column headings.

To format the data correctly, the macro selects the entire `Total Revenue` column and formats it as currency:

```
Range("Table2[Total Revenue]").Select
Selection.NumberFormat = "[$$-en-US]#,##0.00"
```

Finally, the macro ends with the statement `End Sub`.

The macro as defined contains specific references to worksheet 2 (such as the range of cells in the table), so when it is applied to another sheet it does not produce the desired results. To fix this, you must generalize the macro. You can do this manually, or you can use ChatGPT. For best results, use the ChatGPT modality optimized for reasoning (at the time of writing this chapter, this modality is Thinking). To select a specific modality in ChatGPT, log in to the web interface and click the name of the current template at the top left. A drop-down menu opens with all available modality (Figure 4-17). Select the modality for reasoning (Thinking in my case).

Now, ask ChatGPT to modify the VBA code by removing all specific references to the worksheet. Use the following prompt:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
Modify the following macro as follows: replace the specific sheet with the active
sheet, replace the reference to the specific cells A1:B6 with all the nonempty
cells, and don't set the name of the table:
Sub ConvertToTable()
'
' ConvertToTable Macro
' Converts a sheet into a table and formats revenue as currency
'
' Keyboard Shortcut: Ctrl+Shift+T
''
    Range("A2").Select
    Application.CutCopyMode = False
    ActiveSheet.ListObjects.Add(xlSrcRange, Range("$A$1:$B$6"), , xlYes).Name = _
        "Table2"
    Range("Table2[Total Revenue]").Select
    Selection.NumberFormat = "General"
    Selection.NumberFormat = "[$$-en-US]#,##0.00"
End Sub
```

Figure 4-17. The drop-down menu in the ChatGPT web interface allows you to choose the modality to use

ChatGPT returns a new macro code, also commenting on the points it changed. In my case, it produced the following code:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
Sub ConvertToTable()
    '
    ' ConvertToTable Macro
    ' Converts a sheet into a table and formats revenue as currency
    '
    ' Keyboard Shortcut: Ctrl+Shift+T
    '
    Dim ws As Worksheet
    Dim rng As Range

    ' Work with the active sheet
    Set ws = ActiveSheet

    ' Identify the range of all nonempty cells
    Set rng = ws.UsedRange

    ' Clear any copy/paste mode
    Application.CutCopyMode = False

    ' Create a table from the UsedRange. The table will have a default name.
    ws.ListObjects.Add(SourceType:=xlSrcRange, Source:=rng, _
    XlListObjectHasHeaders:=xlYes)

    ' Format column B (2nd column) as currency
    ws.Columns("B:B").NumberFormat = "#,##0.00 $"

End Sub
```

Replace the original code of the ConvertToTable macro inside the VBA Editor with the code generated by ChatGPT. If the VBA Editor does not highlight any lines in red it means that the code does not contain errors. Otherwise, the code generated by ChatGPT is incorrect. In my case, ChatGPT generated a syntax error code, as shown in Figure 4-18.

WARNING

If you replace the entire code, including the `Sub` and `End Sub` lines, you'll need to manually put the keyboard shortcut back in. You'll need to do this every time you replace the code. To restore the shortcut key, select Macros from the Developer tab. Then select the ConvertToTable macro and click Options. A new dialog box opens. Reenter the shortcut key Ctrl + Shift + T and click OK. In the original dialog box, click Run if you want to run the macro or Cancel if you don't want to run it.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 4-18. The error highlighted by the VBA Editor

To resolve the error, continue the conversation with ChatGPT by typing the following prompt:

```
Rewrite and optimize the code by fixing the following error: ' Create a table
from the UsedRange. The table will have a default name.
ws.ListObjects.Add(SourceType:=xlSrcRange, Source:=rng, _
XlListObjectHasHeaders:=xlYes)
```

Replace the error with the specific error you got. ChatGPT generates a new version of the code. In my case, it generated the following code:

```
Sub ConvertToTable()
    '
    ' ConvertToTable Macro
    ' Converts the active sheet's used range into a table
    ' and formats revenue as currency in Column B.
    ' Keyboard Shortcut: Ctrl+Shift+T
    '
    Dim ws As Worksheet
    Dim rng As Range

    ' Work with the active sheet
    Set ws = ActiveSheet

    ' Identify the range of all nonempty cells
    Set rng = ws.UsedRange
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
    ' Clear any copy/paste mode
    Application.CutCopyMode = False

    ' Create a table from the UsedRange.
    ' By using HasHeaders:=xlYes, Excel will try to detect column headers
    ' automatically.
    ' The table will have a default name (e.g., Table1, Table2, etc.).
    ws.ListObjects.Add SourceType:=xlSrcRange, _
                       Source:=rng, _
                       HasHeaders:=xlYes

    ' Format column B as currency
    ws.Columns("B:B").NumberFormat = "#,##0.00 $"
End Sub
```

Paste the code into the VBA Editor. In this case, it contains no syntax errors, so we can analyze it to determine if it contains any conceptual errors.

ChatGPT eliminated the reference to specific cells, selecting all non-empty cells:

```
' Identify the range of all nonempty cells
    Set rng = ws.UsedRange
```

Next, it removed the reference to the specific table name. Conceptually, the macro seems correct, so we can save it and apply it to the Sales Trends worksheet to see if it works:

1. Click the Sales Trends worksheet.
2. Select the Developer tab in the Ribbon and then Macros.
3. Select the ConvertToTable macro and click Run.

In my case, when running the macro, the VBA compiler gives the following error:

```
Compile error: Named argument not found
```

The error is related to this line of code:

```
ws.ListObjects.Add SourceType:=xlSrcRange, _
                   Source:=rng, _
                   HasHeaders:=xlYes
```

Use ChatGPT again to fix the problem. Continue the conversation by writing the following prompt containing the error generated by the compiler:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
I got this error: Compile error: Named argument not found.
```

ChatGPT tries to fix the problem by generating a new version of the code. In my case, it generated the following code:

```
Sub ConvertToTable()
    '
    ' ConvertToTable Macro
    ' Converts the active sheet's used range into a table
    ' and formats revenue as currency in Column B.
    ' Keyboard Shortcut: Ctrl+Shift+T
    '
    Dim ws As Worksheet
    Dim rng As Range

    ' Work with the active sheet
    Set ws = ActiveSheet

    ' Identify the range of all nonempty cells
    Set rng = ws.UsedRange

    ' Clear any copy/paste mode
    Application.CutCopyMode = False

    ' Create a table from the UsedRange using positional arguments:
    '   1. SourceType (xlSrcRange)
    '   2. Source (rng)
    '   3. LinkSource (omit or blank)
    '   4. XlListObjectHasHeaders (xlYes or xlNo)
    ws.ListObjects.Add xlSrcRange, rng, , xlYes

    ' Format column B as currency
    ws.Columns("B:B").NumberFormat = "#,##0.00 $"
End Sub
```

Quickly check the generated code. Since everything looks correct, paste it into the VBA Editor and save it. Now you are ready to apply the macro to a new worksheet. Follow these steps to apply the macro to the Sales Trends worksheet:

1. Select the Sales Trends worksheet.
2. Select Macros from the Developer tab.
3. Select the ConvertToTable macro in the dialog box.
4. Click Run.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

At this point, the data contained in the Sales Trends worksheet is correctly converted into a table and formatted as required. Reapply the macro to the other worksheets (except the Payment Methods worksheet, which does not contain monetary data).

TIP

Use ChatGPT not only to modify the code of an existing macro but also to generate and debug the code.

From the example described we can extract a general workflow (Figure 4-19) that you can always apply to generalize the macros you record.

Figure 4-19. The workflow to generalize the recorded macros

First record the macro, then copy the VBA code of the macro and ask ChatGPT to generalize the code, eliminating all parts that specifically refer to a particular worksheet. Quickly check if the code generated by ChatGPT is correct and test it in Excel. If there are problems, use ChatGPT to troubleshoot. Finally, run the macro.

You can further generalize the workflow by directly asking ChatGPT to generate the macro in VBA, without recording it manually. Let's see how to proceed in the next section.

Using ChatGPT for Macro Generation

In previous chapters, you have seen the importance of defining the context to ensure that ChatGPT generates the correct result. You also saw that you can define the context in two ways:

1. Describe the structure of your dataset in words, specifying the name of the columns and possibly also the name of the Excel table, if the dataset is formatted as an Excel table.
2. Provide ChatGPT with the dataset (or a portion thereof) you want to operate on as input directly. This second option is faster but is less suitable if you are dealing with sensitive data.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Even if you want to have ChatGPT generate macro code directly, you need to specify the context. As in the other cases, organize your conversation with ChatGPT in the form of a story: context, main part, and action. The context contains the structure of your table, the main part the instructions for the macro, and the conclusions where you want to save the results of your macro.

Let's take an example. Consider the *04/sales_dataset_q5.xlsx* dataset again. Suppose you want to add a new worksheet to your workbook containing a weekly sales report for each product. Use the following prompt with context, main body, and action to ask ChatGPT to generate a macro that creates the report:

```
Consider an Excel table named MainTable representing sales, with the following
columns:
  • Transaction ID
  • Date of Sale
  • Product ID
  • Product Name
  • Quantity Sold
  • Unit Price
  • Total Sale Amount
  • Payment Method
  • Store Location
  • Customer Type
Write a VBA macro to analyze the dataset and generate a report on weekly sales
for each product.
  1. Group by Week and Product: Aggregate the data based on weeks and products,
     calculating the total sales.
  2. Create a New Report Sheet: Save the results in a new sheet named "Sales
     Report".
  3. Format the Report: Convert to a Table.
```

Use ChatGPT's advanced reasoning model (in my case, o1). ChatGPT generates the macro code and a short comment. In my case, ChatGPT generated quite an extensive macro, which you can find in the book's GitHub repository at *04/create_weekly_sales_report_macro.txt*. The VBA code analyzes the sales data in a list (`ListObject`) called `MainTable` in Excel, grouping them by week and by product. It first creates or overwrites a new sheet called Sales Report. Then, it uses a special data structure called a dictionary (`Scripting.Dictionary`) to add the quantity sold and the total amount of each product, divided by week (calculated with the function `WeekNum`). Finally, it writes the results to the Sales Report sheet and converts them into a formatted table.

Add the macro to the Excel file as follows:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

1. Click the Developer tab in the Ribbon, and in the Code group click Visual Basic.
2. In the Visual Basic Editor, look for the Project pane on the left (it lists all open workbooks and their objects). Right-click VBAProject (*sales_dataset_q5.xlsx*).
3. Select Insert → Module. A new blank module will appear in the code window (usually titled Module1 by default).
4. Paste the code generated by ChatGPT. Click Save if the file is already in the *.xlsm* format. Otherwise, click Save As, give a name to your file, and choose Excel Macro Enabled Workbook as the file format. Finally, close the VBA window.

Run the macro by selecting Macros in the Developer tab and then selecting Run. A new worksheet named Sales Report is added to the workbook, as shown in Figure 4-20.

Figure 4-20. The result obtained by applying the macro

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Exercise 2

Using the *04/sales_dataset_q5.xlsx* dataset, write a prompt that solves the following problem: write a VBA macro that analyzes the dataset and automatically highlights high and low sales based on a user-defined threshold.

You can find the solution in the book's GitHub repository at *04/exercise3_prompt.txt*.

Summary

In this chapter, you learned how to use ChatGPT to analyze a dataset, define macros, and write VBA code:

ChatGPT helps you speed up data analysis in Excel by automatically identifying dataset categories, suggesting relevant questions, and generating advanced formulas to extract insights more effectively.

Following a structured workflow, you can load a dataset, obtain key questions, generate answers via Excel formulas, and implement them directly in the file, automating complex operations without advanced programming knowledge.

Using ChatGPT for VBA macros allows you to create, generalize, and fix code efficiently, reducing errors and making repetitive tasks faster and more customizable to specific spreadsheet needs.

Through automation with macros and VBA, you can generate weekly reports, format data dynamically, and highlight key information, improving data organization and analysis with a more structured and automated approach.

Activity

Solve the crossword puzzle shown in Figure 4-21. The solution is in Appendix D.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 4-21. Crossword to solve

Across

1 - A type of table in Excel used to summarize and analyze data dynamically

4 - Raw facts and numbers that can be analyzed or processed in a spreadsheet

6 - Technology that enables machines to simulate human intelligence and decision-making

7 - Widely used spreadsheet software developed by Microsoft

Down

2 - A programming language used in Excel to automate tasks and create macros

3 - Transactions where goods or services are exchanged for money

5 - A structured arrangement of data in rows and columns within a spreadsheet

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Part II. Data Science with Copilot

In Part II, you'll learn how to use Copilot, the native Microsoft Excel AI tool, to perform the most common data science tasks. Compared to ChatGPT, seen in Part I, Copilot is directly integrated into Excel, thus you can use it directly in your worksheet without uploading the file into an external tool.

Chapter 5 introduces Copilot in Excel to create, edit, and understand your data through natural language prompts. You'll generate new elements such as PivotTables and charts, refine your data by renaming and splitting columns, and use built-in insight features to uncover patterns, summaries, and visualizations. Along the way, you'll see how Copilot supports your workflow step-by-step, helping you move from raw tables to clearer structure and initial data understanding.

In Chapter 6, you'll learn how to use Copilot to prepare your data before analysis by cleaning duplicates, handling missing values, and fixing formatting issues. You'll see how to detect and remove repeated records, choose appropriate strategies to fill or manage gaps in your data, and standardize numbers and dates so they behave correctly in formulas and reports. Along the way, you'll practice writing effective prompts and combining Copilot's suggestions with Excel's built-in tools to turn messy, inconsistent datasets into reliable, analysis-ready information.

In Chapter 7, you'll learn how to use Copilot to understand, create, and refine text formulas in Excel. You'll see how it can explain complex logic in plain language, generate formulas from natural language requests, and adjust them to handle messy, real-world data. You'll also apply conditional text logic to categorize and label information based on keywords or patterns, turning simple descriptions into practical, reliable spreadsheet solutions.

Chapter 8 explains how to use Copilot to calculate, visualize, and interpret basic statistics in Excel. You'll explore measures of central tendency, dispersion, and distribution shape, generating formulas and summary tables directly from natural language prompts. Along the way, you'll create charts such as scatter plots, histograms, and box plots to reveal patterns that numbers alone might hide, and you'll use Copilot to

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

help interpret results, assess reliability, and spot skewness, outliers, or dominant values.

In Chapter 9, you'll use Copilot to perform more advanced data analysis in Excel. You'll create dynamic PivotTables, add interactive filtering, and structure effective prompts to guide Copilot's help. You'll also run What-If Analysis with tools like Goal Seek and data tables to explore how changes in key inputs affect outcomes. Finally, you'll learn how to ask Copilot to suggest relevant analyses and why it's essential to review and validate its reasoning instead of accepting results blindly.

Chapter 10 explains how to use Copilot to transform Excel data into clear visualizations and structured reports. You'll see how to design effective charts by choosing the right visualization type; structuring prompts with context, main point, and action; and simplifying visuals to remove clutter. Beyond individual charts, you'll build complete reports with summary pages, raw data descriptions, key statistics, and organized worksheets. Finally, you'll create an interactive dashboard by combining charts and using Copilot to help automate layout and controls, turning analysis into polished, audience-ready reporting.

Appendix B in Part IV continues the short novel introduced in the first chapter, this time applying the concepts covered in Part II. As you move through the material, you'll complete hands-on activities woven into a brief adventure narrative, combining practical learning with a story-driven experience.

Chapter 5. Getting Started with Copilot

One morning, after breakfast, my children and I gathered a few leftover cookie crumbs and brought them out to the garden, near a spot where we had often seen ants passing by. We placed the crumbs down one at a time, deliberately, curious to see what would happen. Within minutes, the first ants began to appear. Quietly and methodically, each one took on a task: some collected, others carried, and a few seemed to organize the flow. Not a single crumb was wasted. In no time, everything had been neatly picked up and taken away, with remarkable precision. That simple moment reminded me of how Copilot in Excel works. It doesn't require a grand gesture or a perfect setup. Instead, it responds best when you offer it a little at a time. A question. A column. A hint of what you're trying to achieve. You don't need to know everything up front: you just need to start.

Copilot gathers the pieces you provide, connects them, analyzes them, and builds valuable insights, step-by-step. One crumb at a time, it helps sort, visualize, and make sense of data. It doesn't replace your judgment, but it frees you from repetitive tasks so you can focus on what matters most: understanding and decision-making.

In this chapter, we'll explore how to get started with Copilot in Excel, from enabling it in your environment to using its first key features for creating, editing, and understanding your data in an intuitive and guided way. To begin, you don't need much, just a willingness to offer the first crumb. More specifically, this chapter covers the following:

- Introducing and enabling Copilot in Excel
- Creating
- Editing
- Understanding

Let's start with introducing and enabling Copilot in Excel.

Introducing and Enabling Copilot in Excel

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Microsoft Copilot in Excel is an AI-driven assistant for managing spreadsheets. You can use Copilot to analyze data, suggest formulas, automate repetitive tasks, and build visualizations.

Before using Copilot in Excel, make sure your Microsoft 365 subscription and Copilot access are set up correctly: Microsoft offers a free basic Copilot experience that you can use for general AI assistance on the web, in Windows, or in the Copilot app without a paid license, but this free version does not include the full AI integration inside Excel and other Microsoft 365 applications. For the advanced AI features in Excel, you need a paid Copilot license or an appropriate Microsoft 365 plan that includes Copilot (such as Microsoft 365 Premium for individuals or a Copilot add-on for business/enterprise subscriptions).

Once you have access to Microsoft Copilot, the Copilot icon should appear directly in the Ribbon's Home tab. To check if it is enabled, launch Excel and verify the presence of the icon, as shown in Figure 5-1.

Figure 5-1. The Copilot icon on the Ribbon's Home tab

Click the Copilot icon to access the Copilot panel (Figure 5-2). Note that Copilot works only on files saved on OneDrive.

As of February 2026, the default Microsoft Copilot chat experience functions as a conversational assistant. It provides foundational help such as explaining formulas, offering step-by-step instructions for completing tasks, and answering general "how-to" questions. These core capabilities are available with a paid Microsoft subscription.

In addition, the paid version of Copilot expands on this baseline experience by introducing two additional operating modes:

Agent mode (Frontier)

> Performs tasks directly in your workbook, such as creating new worksheets and PivotTables, cleaning data, and so on.

Analyst

> Calculates statistics and general metrics based on your data. The original worksheet is not modified.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 5-2. The Copilot panel

You can access these modes through the Tools button in the Copilot pane's text box, enabling more advanced and task-oriented interactions beyond standard chat assistance.

Because the Copilot interface evolves frequently, this book uses the general term *Copilot* when describing how to complete tasks. Rather than focusing on the specific

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

labels used in the interface at any given time, we focus on the type of task you want to accomplish. As a practical guideline, use Agent mode when you need Copilot to modify the original worksheet or take direct action on your content. Use Analyst mode when your goal is to perform calculations, analyze data, or generate insights. Even if the interface changes in future versions, this distinction between action-oriented tasks and analysis-oriented tasks remains consistent.

WARNING

The first time you launch Copilot, it might display the message "I only work with files that have AutoSave turned on. Turn it on to continue." It then prompts you to add the file to OneDrive.

The Copilot panel offers various options, which, at the time of writing, can vary from run to run. Possible options include "How do I?," "Suggest a column," "Add a color scale to a column," "Split columns," and so on. Click one of the proposed options to access the related feature. Alternatively, write your question directly in the chat box.

Now that your Copilot assistant is installed and working, let's use it to perform some basic operations, including creating, editing, and understanding:

Creating

Generate new content, such as PivotTables or charts

Editing

Modify or refine existing data, formulas, or tables

Understanding

Provide insights, summaries, and explanations to help interpret information

The create, edit, and understand operations encompass the most common methods for interacting with your data. Let's dive in and see what these categories can help you do. We will use the dataset "Persons reading books in the last 12 months by sex, age, educational attainment and number of books (ilc_scp27)" provided by Eurostat under the CC BY 4.0 license and available in the GitHub repository as *05/reading_habits.xlsx*.

Table 5-1 shows an extract of the dataset.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 5-1. An extract of the reading habits dataset

freq	n_book	isced11	sex	age
A	0	ED0-2	F	Y16-24
A	0	ED0-2	F	Y16-24
A	0	ED0-2	F	Y16-24
A	0	ED0-2	F	Y16-24
A	0	ED0-2	F	Y16-24
A	0	ED0-2	F	Y16-24
A	0	ED0-2	F	Y16-24
A	0	ED0-2	F	Y16-24
A	0	ED0-2	F	Y16-24
A	0	ED0-2	F	Y16-24

The dataset provided by Eurostat refers to the number of people who read books in the past 12 months, broken down by various demographic and educational characteristics. The dataset contains the following columns:

`Freq`

: Frequency of the data collection. In this case, it is always A, corresponding to annual data.

`n_book`

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Number of books read. Possible values include 0, LT5 (less than 5), 5-9, and GE10 (greater than or equal to 10).

`isced11`

Level of education according to the ISCED 2011 classification. Possible values include: ED0-2 (low education level, e.g., primary school or less), ED3-4 (Secondary education), ED5-8 (Tertiary education, e.g., university or postgraduate).

`Sex`

Sex. Possible values include F (female), M (male), T (total).

`Age`

Age group. Examples of values include Y16-24 (16–24 years old), Y25-34 (25–34 years old), and Y_GE65 (65 years and over).

`Unit`

Unit of measurement. PC stands for percentage.

`geo\TIME_PERIOD`

Country or geographical area. For example: AT: Austria, BE: Belgium, EU27_2020: European Union (27 countries, as of 2020).

`2022`

percentage of people who read books in the past 12 months in 2022, based on demographic and educational characteristics.

You are now ready to use Copilot with this dataset. Let's start by creating something new with Copilot. For all the examples, I'll use the web version of Excel with only the Copilot subscription active (no Microsoft 365 subscription). However, if you use a different version, you should not experience too many differences.

Creating

When you use Copilot to create, you're making something new, such as a PivotTable. In this section, we'll focus on creating PivotTables and column charts.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

PivotTable

Consider the reading habits dataset. We'll use Copilot to quickly generate a PivotTable that shows how book-reading behavior varies across different education levels. Proceed as follows:

1. Open the reading habits dataset available on GitHub (*05/reading_habits.xlsx*) in Excel. The dataset consists of a single sheet with a single table. Click anywhere in the table (such as cell A2).

2. Click the Copilot button in the Ribbon. In the chat box, write the following prompt in natural language and then click the Send button or press the Enter/Return key:

   ```
   Consider the table Table1. Create a PivotTable that shows the average
   percentage of people who read books, with education level (isced11)
   as rows and the number of books read (n_book) as columns.
   ```

   Figure 5-3 shows a portion of the output generated by Copilot in my case.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 5-3. The output generated by Copilot after being asked to generate a PivotTable

3. In this example, the answer generated by Copilot is structured in three parts:

    a. A description of the PivotTable structure (not shown in Figure 5-3), including rows, columns, and calculated values.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

b. A table with the key insights extracted from the data, as shown in Figure 5-3.

4. Copilot has already added the PivotTable to the worksheet as a new spreadsheet, named PivotTable1. Figure 5-4 shows the resulting PivotTable.

Figure 5-4. The PivotTable generated by Copilot

Reviewing the generated output, Copilot has correctly calculated the average value for each column except for the `Grand Total`. This column should calculate the total percentage and not the average value. Since the `Grand Total` value should always be 100%, we can remove this column. You have to proceed manually because Copilot can't currently edit PivotTables for you. Select the cell corresponding to `Grand Total` (e.g., cell F3) and right click it. Select Remove Grand Total (Figure 5-5).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 5-5. The procedure to remove the `Grand Total` column

Column Chart

Let's now use Copilot to create a column chart showing the reading habits for ED3-4 (secondary education). To draw the column chart correctly, we must prepare the data before asking Copilot to generate the chart.

Follow these steps to have Copilot generate a new column chart for ED3-4:

1. Filter the data in the PivotTable to display only information related to ED3-4. Right-click the drop-down arrow near the `Row Labels` header and from the drop-down menu, keep only ED3-4 selected by deselecting the other options (Figure 5-6).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 5-6. The drop-down menu for the `isced11` column

2. The filter is now enabled and identified through a small symbol near the `Row Labels` header (Figure 5-7).

Figure 5-7. The PivotTable with the filter enabled for the `isced11` column

3. To build the column chart, columns B, C, D, and E must appear as rows of a table; otherwise, the chart will draw them as different series. For this reason, we must transpose the table and convert columns to rows:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

a. Select all the PivotTable values manually, or click Ctrl+A (or Cmd-A for macOS).

b. Right-click the selected items and click Copy from the drop-down menu. Alternatively, click Ctrl+C.

c. Add a new sheet by clicking the + button near the current sheet name. Move to the new sheet.

d. Open the Paste menu from the Home tab of the Ribbon, and in the drop-down menu, click Transpose Rows and Columns (or simply Transpose). Figure 5-8 shows the resulting range.

Figure 5-8. The transposed table

4. Click cell A1. Now you are ready to ask Copilot to draw the column chart. Write the following prompt in the Copilot chat box:

```
Draw a column chart ordered by decreasing ED3-4. Use the labels in the
Row Labels column for the x-axis.
```

Figure 5-9 shows the output generated by Copilot.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 5-9. The output generated by Copilot

5. Copilot has also generated a new sheet containing the resulting column chart. Figure 5-10 shows the final result, featuring a new PivotTable on the left and a column chart on the right side of the sheet.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

WARNING

Sometimes, Copilot may not generate the correct answer immediately, but it could provide an answer similar to the following one:

```
I encountered an issue while trying to draw the column
chart. You can manually create a column chart by
selecting the data in the pivot table and using the
"Insert" tab to choose a column chart type.
```

Copilot could also generate an answer similar to the following one:

```
Something went wrong. Please try again in a moment.
```

You can use the thumbs up and thumbs down button to give feedback on what went wrong and then simply rewrite the prompt to solve this problem and make Copilot generate the correct answer.

Figure 5-10. The PivotTable and column chart generated by Copilot

Exercise 1

Consider again the reading habits dataset. Write the prompt for Copilot to create a PivotTable showing the average percentage of people who read books, with age group as rows and the number of books read as columns. Figure 5-11 shows the resulting PivotTable.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

You can find a possible prompt for Copilot in GitHub under *05/exercise1_solution.txt*.

TIP

Just like a vine that quietly grows and, over time, produces clusters of grapes, the process of creating with Copilot begins with something small and simple. You plant an idea, such as a prompt, a question, or a need, and let the system guide its growth. From that single start, you can generate insights, visualizations, or new ways of looking at your data that you hadn't considered before. The beauty lies not in complexity, but in the natural flow of creation that Copilot supports.

Now that we've explored what it means to create with Copilot, let's move to editing.

Figure 5-11. The PivotTable from Exercise 1

Editing

In this section, you will see how to perform basic editing operations on a spreadsheet using Copilot: renaming column headers and splitting a column into two. These basic operations are meant to whet your appetite for learning more about how to perform other operations using Copilot. Later in the book, you will encounter additional operations, including data cleaning and preparation.

Let's start with renaming columns.

Renaming a Column

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

I have often worked with datasets that have columns with difficult names to remember, such as acronyms or code names. Renaming these columns helped me work better. Sometimes, however, I couldn't find the right name for the columns. At other times, I didn't have time to think of sensible names, so I continued to work with acronyms and incomprehensible names, which increased the processing and analysis time.

To solve this problem, I found great help in Copilot. Consider the dataset related to reading habits once again. Select a cell in the original table, start Copilot, and type the following prompt in the Copilot chat box:

```
Propose alternative column headers so that they reflect the column content.
```

Copilot generates alternative column headers, like those shown in Figure 5-12.

Figure 5-12. The alternative column names proposed by Copilot

Unfortunately, at the time of writing, Copilot cannot automatically replace the generated column names with the proposed ones. This operation must be done manually. So, you can replace the existing column names with the proposed column names by simply

writing the proposed column names in the corresponding column headers. Figure 5-13 shows the resulting table.

Figure 5-13. The resulting PivotTable

TIP

You can also use this technique to translate column names into another language. For example, use the following prompt:

Propose a translation into Italian of the column names.

Copilot then provides suggested translations of the column headers into Italian:

```
Frequency: Frequenza
Book Count: Conteggio dei libri
Education Level: Livello di istruzione
Gender: Genere
Age Group: Gruppo di età
Unit: Unità
Geographical Area: Area geografica
Value for 2022: Valore per il 2022
First age: Prima età
Last age: Ultima età
```

Splitting a Column

Splitting means dividing a column into two or more columns based on one or more criteria. For example, you can split a column containing the first and last name of a person into two columns, one for the first name and one for the last name. Considering again the dataset on reading habits, select the cell in the original table (where you renamed the column headers), and type the following prompt in the chat box:

```
Extract the first and last age in the Age Group column and insert them into two
new columns.
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

In this example, the prompt asks Copilot to split the `Age Group` column into two columns, one for the first age and one for the last age. As a result, Copilot generated two outputs, one for each new column:

```
Extracts the first age from the age group by taking the part before the hyphen.
This helps in identifying the starting age of each age group in the table.
=TEXTBEFORE([@[Age Group]],"-")

Extracts the last age from the "Age Group" column by taking the text after the
hyphen. This helps to identify the upper limit of the age range for each entry
in the table.
=TEXTAFTER([@[Age Group]],"-")
```

Each explanation is accompanied by the specific formula to use to extract the text to insert into the new column.

TIP

Next to each formula, Copilot also displays the "Show explanation" option (Figure 5-14), which displays details about the formula.

Figure 5-14. The "Show explanation" option

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

TIP

For example, clicking the "Show explanation" option below the formula for the `Last Age` column produces output similar to this:

```
Extracts the last age from the "Age Group" column in the
table by using the following steps:
Uses the TEXTAFTER function to find the text that appears
after the hyphen ("-") in the "Age Group" column value.
```

In addition to the formulas that generate the new columns, Copilot also displays a preview of the result columns in the chat box (Figure 5-15).

Figure 5-15. A preview of the calculated columns in the Copilot chat box

The first column contains a slight inaccuracy: the letter Y (which stands for Year) is before the actual age. This value is not necessary; ask Copilot to remove it:

```
Modify the following formula to exclude the starting
Y: =TEXTBEFORE([@[Age Group]],"-") and recalculate
the 2 new columns to insert
```

Copilot responds by modifying the original formula. In my case, it generated the following output:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
=TEXTBEFORE(TEXTAFTER([@[Age Group]],"Y"),"-")
```

In addition to changing the formula, Copilot once again lets you preview the table (Figure 5-16), this time without the leading Y in the `First age` column.

Figure 5-16. A preview of the modified table

Now, click the "+ Insert columns" button below the table preview in the Copilot chat box. Copilot automatically adds the new columns to the original table, as shown in Figure 5-17.

Figure 5-17. The resulting table with the two new columns

Exercise 2

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Write the prompt to extract the first and last level in the `Education Level` column and insert them into two new columns. If the Education Level is TOTAL, ask Copilot to write 0 in the `First level` column and 8 in the `Last level` column. Figure 5-18 shows the resulting table.

Figure 5-18. The table for Exercise 2

You can find a possible prompt for Copilot in GitHub under *05/exercise2_solution.txt*. If Copilot only calculates the `First level` column, type this additional prompt:

```
Now add the Last level column.
```

TIP

To understand what editing data is, consider water. Water's essence doesn't change, but when poured into a glass, a bottle, or a riverbed, it takes on a different shape each time. The same is true for data. The raw content remains the same, but when you edit and reshape it, by renaming, splitting, or organizing columns, you begin to see it in a new form. Editing is about preparing the container, setting the stage for meaning to emerge. After preparing the container, it's time to go deeper. In the next section, you'll explore how to use Copilot to understand the data—not just what it says, but what it means.

Understanding

Once your data looks good, it's time to ask the real questions. Who reads more books, men or women? What age group reads the most? How has reading changed over time? In this section, you'll see the default strategy provided by Copilot to understand data. In

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

the following chapters, you'll delve into more advanced techniques that will help you to extract meaning from data.

When you launch Copilot from the Home tab of the Ribbon, Copilot provides some default options you can use to quickly explore your data. Among the default options is "Show data insights." If you don't see this option immediately, click "More examples" in the Copilot chat box until this option appears. This option also appears in the chat box while you are working on a specific table, as shown in Figure 5-19. If you don't see "More examples," click the Refresh button next to the suggestions.

Figure 5-19. The "Show data insights" button in the Copilot chat box

Let's use an example to show how this option works:

1. Consider the reading habits dataset again and select the worksheet where you added the `First age` and `Last age` columns.
2. Click the "Show data insights" button in Copilot. Copilot will calculate the frequency of values for a specific column, as shown in Figure 5-20.

   Additionally, Copilot generates a text summary of relevant information related to the calculated insight. In my case, Copilot wrote the following text:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
The data insights show the count of 'Frequency' by 'Last age'. For example,
the count for 'Last age' 64 is 6528, and for age '54' is 3264. See the
table above for more details.
```

Figure 5-20. The chart generated by Copilot when clicking the "Show data insights" button

3. Click the "Add to a new sheet" button to add the calculated insight to a new worksheet. Figure 5-21 shows the resulting sheet, with the generated PivotTable and the bar chart.

TIP

An *insight* is a valuable piece of information extracted from data. An insight differs from data in that data is a raw piece of information collected from a source, whereas an insight is the meaning extracted from the data. For example, 3 million followers is data; 3 million more followers than last year is an insight.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 5-21. The new worksheet generated by Copilot, containing the extracted insight

Once you have completed the elaboration of an insight, Copilot offers two new buttons: “Can I see another insight?” and “Add all insights to grid,” as shown in Figure 5-22. If you don’t have the original data selected, you might see the buttons “Go back to data” and “Undo.” Click the first button to go back to the table data.

Figure 5-22. The two buttons provided by Copilot after elaborating on the first insight

If you click the “Can I see another insight?” button, Copilot will follow the same procedure as in the previous case and will calculate another insight. In my case, it calculated the frequency by first age and gender, as shown in Figure 5-23.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 5-23. The result of clicking "Can I see another insight?"

Click "Add to a new sheet" if you want to add this insight to your workbook.

WARNING

When using the web version of Excel, your browser settings might prevent some operations required by Copilot. In this case, a message similar to the one shown in Figure 5-24 can appear in the Copilot chat box. To solve the problem, simply click the Verify button in the Copilot chat box.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 5-24. The Verify button in the Copilot chat box

Exercise 3

Select the first worksheet of the reading habits dataset and click the "Add all insights to grid" button in Copilot (Figure 5-22). Describe what Copilot suggests. You should obtain something similar to Figure 5-25.

Figure 5-25. A possible output after clicking the "Add all insights to grid" button in the Copilot chat box

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

TIP

In some cases, the "Add all insights to grid" button doesn't appear. If it doesn't appear, follow these steps:

1. Type the prompt `Add all insights to grid.` Copilot usually tells you it can't do that but shows the suggestion "Add all insights to grid."
2. Click on the suggestion to add the insights.

In this particular example, the insights provided by Copilot's default analysis are not especially impressive. However, in other situations, they can serve as a helpful starting point for exploring a dataset. Starting in the next chapter, you'll see more advanced techniques to use Copilot to prepare and analyze data, starting with data preparation.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

YES, YOU CAN CATCH THE BUTTERFLY

We've reached the end of this introductory chapter on Copilot, and depending on how things went for you, you might be feeling either excited or a bit discouraged. Honestly, over the past few days, I've felt a bit disheartened while using Copilot, and I'd like to share with you why.

As I write this, I'm tucked away in the mountains, spending time with my family. It's peaceful and beautiful, but the internet connection? Not so great. It's slow, and sometimes cuts out completely. And here's the thing: Copilot requires a stable internet connection to function correctly. In my current situation, I have to retry commands multiple times before getting any helpful response. It's not Copilot's fault; it's the connection. But still, it's frustrating. And yeah, there have been moments when I've felt like giving up. You might be feeling the same. Maybe you're thinking, "Ugh, let's just go back to the old ways and toss this book (and Copilot) out the window."

But before you do, let me tell you a little story. Yesterday, my son Antonio was running through a meadow, chasing a stunning, rare blue butterfly. I asked him, "What are you doing, Antonio?" And he said, "I'm going to catch it!" I laughed and told him, "That's impossible!" But he looked at me and said, "No, I can do it."

I watched him: again and again, he ran, chased, reached, and missed. All of a sudden, he sprinted toward me, beaming: "Mom, look!" In his hand was the butterfly, fluttering. He showed it to me with pride, and gently placed it on a flower and let it go.

What a powerful lesson from a six-year-old.

The same applies to Copilot and the use of generative AI in general: don't give up at the first obstacle. Keep going. Keep trying. Be curious. Be persistent. And I promise, the magic will start to happen, little by little.

Summary

In this chapter, you learned how to use Microsoft Copilot within Excel for three simple tasks: creating, editing, and understanding:

> Copilot helps you simplify data tasks in Excel by allowing you to interact with spreadsheets through natural language. Whether you're generating a

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

PivotTable, building a chart, or exploring your data, Copilot guides you through the process of creating, editing, and understanding data, and suggests relevant actions.

Use Copilot to create PivotTables and visualizations such as column charts, and to suggest or translate column names.

Copilot can analyze your table and offer key insights, including charts and summaries. With options like "Show data insights" and "Add all insights to grid," you can explore your data with just a few clicks and uncover patterns effortlessly.

While Copilot still has some limitations (for example, the inability to rename headers or potential issues with browser compatibility), its intuitive interface and ability to answer natural language prompts make it a valuable partner for data exploration and automation. Copilot is constantly improving. Features that weren't there in the early days are now available.

In this chapter, you got your first taste of what Copilot can do. But we're just getting started. In the next chapter, you'll learn how to harness Copilot to prepare your data like a pro, including removing duplicates, handling missing values, and applying clean, consistent formatting. Get ready to unlock the full potential of your dataset.

Activity

Solve the crossword puzzle in Figure 5-26. After completing it, you'll read the word used to indicate something valuable extracted from data. The solution is in Appendix D.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 5-26. Crossword puzzle to solve

Clues:

1. The process of modifying existing data within a worksheet, such as renaming columns, splitting cells, or cleaning up labels
2. The process of interpreting data with the help of Copilot's analysis features, like "Show data insights" (involves generating summaries, visualizations, and key patterns that help users make sense of the information)
3. A formatting function often used before creating charts, where the orientation of data is switched from rows to columns or vice versa
4. A specific editing task where Copilot is asked to divide a single column (for example, age range or education level) into multiple columns using functions like `TEXTBEFORE()` and `TEXTAFTER()`, thereby helping to structure the data more effectively
5. A visual representation of data used to show relationships, trends, or comparisons between values
6. The title of a column in a dataset
7. The process of using Copilot to generate new elements in a worksheet, such as PivotTables or column charts, based on user prompts

Chapter 6. Data Preparation with Copilot

Many years ago, I found myself tackling a predictive modeling project for the very first time. The goal was to forecast ship routes as part of a project funded by the European Space Agency (ESA). It was also my first real encounter with massive datasets concerning historical ship route data. The dataset comprised approximately 80 million records, provided by two different providers. At the time, that number felt enormous to me. This was also my first hands-on experience building a real-world machine learning model.

I still vividly remember the first version of the model I built. I trained the model using approximately 20 million records and then tested it on the remaining 60 million records. The result was a complete disaster.

First off, the training process took over a month. But that wasn't the worst part; the model's predictions were totally off. It just couldn't get the ship routes right. But I didn't give up. I decided to take a deep dive into the data, examining it in fine detail to gain a deeper understanding of what was really going on. That's when everything started to make sense. I realized that the two data providers had given me almost identical data, so nearly every record was duplicated. In other words, I didn't actually have 80 million records. I had more like 40 million. Then I noticed other issues: incorrectly formatted numbers, missing values, fields that had been interpreted as text instead of numeric values…the list went on.

So, I rolled up my sleeves and got to work, meticulously cleaning and preparing the dataset. After all the cleaning, I was left with around 4 million solid records. I used those to train a new version of the model, and this time, it actually worked. It became fairly accurate at predicting ship routes.

That experience taught me a valuable lesson: before diving into any kind of data analysis, the data needs to be properly prepared and cleaned. And that's exactly what this chapter is all about: getting your data ready for action.

More specifically, this chapter covers three techniques for data preparation:

- Removing duplicates

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

- Dealing with missing values
- Data formatting

Let's start with the first step: removing duplicates.

Removing Duplicates

Duplicates in a dataset are records that appear more than once and contain identical values across all the columns. These repetitions can occur due to manual data entry errors, system imports, or inconsistent data collection processes.

While duplicates may seem harmless at first glance, they can significantly degrade the quality and reliability of your data analysis because they can do the following:

Skew summary statistics and totals

> Duplicates can artificially inflate metrics such as counts, sums, averages, and percentages. For example, if a Sale ID appears twice by mistake, total revenue figures may be overestimated, and average sales per transaction may appear higher than they actually are.

Lead to inaccurate insights

> Duplicates can distort trends, introduce noise into predictive models, and bias results. For instance, customer segmentation based on transaction frequency may incorrectly categorize clients if duplicate entries exaggerate their activity.

Complicate data analysis

> Analysts often rely on the uniqueness of certain identifiers (like Sale ID) to join tables or group data. Duplicates in these fields can lead to incorrect joins, mismatches, or ambiguous aggregations.

Therefore, you should remove duplicates to maintain the integrity of your dataset and the trustworthiness of your results. Let's see how Copilot can help with this task.

Consider the dataset related to sample product sales, which is available in the book's GitHub repository at *06/product_sales.xlsx* (Figure 6-1).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 6-1. The product sales dataset

Each row describes a transaction. The columns of the dataset are as follows:

`Sale ID`

Unique identifier for the sales transaction (e.g., 1001, 1002, and so on)

`Date`

Date of the transaction

`Product Name`

Name of the product sold (e.g., Laptop, Mouse, Monitor)

`Quantity`

Number of units of the product sold in the transaction

`Price`

Unit price of the product

`Total`

Total amount of the sale

`Salesperson`

Name of the salesperson who made the sale

`Branch Location`

Branch where the sale occurred (e.g., New York, Chicago, Los Angeles)

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Before you start to use Copilot, I want to remind you of a few pieces of advice (applicable as of February 2026):

- Use Agent mode if you plan to modify the original worksheet, for example, by adding columns or rows to a table
- Use Analysis mode if you plan to calculate metrics or statistics that you do not want to include in the worksheet
- Use the simple chat for general questions and advice.

Throughout this chapter, I won't prescribe a specific modality. Instead, I'll leave room for you to determine which approach best fits your scenario. This choice is intentional. The Copilot interface evolves continuously, and focusing on the exact name of a feature today may only create confusion tomorrow. Rather than anchoring you to labels that may change, this chapter emphasizes underlying concepts. Once you understand the principles, you'll be able to apply them confidently, no matter how the Copilot interface evolves in the future.

Let's use Copilot to check if the dataset contains duplicates. Open the Copilot chat by clicking its button in the Ribbon, as explained in the previous chapter. Then write the following prompt in the chat:

```
Highlight duplicates in the column Sale ID
```

In most cases, Copilot should provide an answer similar to the one shown in Figure 6-2.

Figure 6-2. The answer provided by Copilot after asking it to highlight duplicates

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

In practice, Copilot asks to apply a conditional formatting rule for duplicate records. Click the Apply button in the Copilot chat. Figure 6-3 shows the result after clicking the Apply button.

Figure 6-3. The dataset with the highlighted duplicate rows for the `Sale ID` column

The dataset contains just one duplicate row. Let's ask Copilot to remove it through the following prompt:

```
Remove duplicate entries
```

At the time of writing, Copilot can't remove duplicates directly. Instead, it suggests the steps to remove duplicates using the tools provided by Excel. In my case, Copilot provided the following answer:

```
I can't help with this.
To remove duplicate entries in the "Sale ID" column of the table "Table1," follow
these steps:
  1. Select the entire table or the "Sale ID" column.
  2. Go to the Data tab on the ribbon.
  3. Click Remove Duplicates in the "Data Tools" group.
  4. In the dialog box, ensure "Sale ID" is checked (or any other columns you
     want to consider for duplicates).
  5. Click OK to remove duplicates.
This will remove rows with duplicate "Sale ID" values while keeping the first
occurrence.
```

Follow the steps proposed by Copilot. Specifically, when you click Remove Duplicates in the Data Tools group, the dialog box shown in Figure 6-4 appears. Select only Sale ID and click OK.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 6-4. The dialog box to remove duplicates

After clicking the OK button, a window appears with some statistics: “1 duplicate value found and removed; 9 unique values remain.” Click OK. Figure 6-5 shows the resulting table.

Figure 6-5. The product sales dataset, without duplicate rows

TIP

Always specify the column you want Copilot to search for duplicates; otherwise, Copilot may not respond correctly.

Now that you have learned how to remove duplicate records, let’s move on to the next task of data preparation: dealing with missing values.

Dealing with Missing Values

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Missing values occur when no data is stored in a cell or field in a dataset where a value is expected. In Excel, a missing value may appear as a blank cell, a placeholder like N/A, or an error code such as `#VALUE!`.

Missing values can depend on different factors, including but not limited to these:

Data not collected at the time of entry

> Certain fields may not have been filled out because the information was unavailable at the time of data capture. For example, a customer survey might include optional fields such as email or age, which some respondents choose not to provide.

Manual input errors or omissions

> A user entering information into a spreadsheet may accidentally skip a cell or fail to update a record after a change. These omissions are often unpredictable and can be difficult to trace, especially in large datasets.

System failures or interrupted data imports

> Technical glitches, such as network interruptions, software crashes, or broken data pipelines, can lead to partial imports. For example, during a batch upload, the system may crash halfway through, leaving some cells blank in some records.

Mismatched merges from multiple sources

> It's common to encounter join mismatches when combining datasets from different sources, especially when key identifiers are missing or inconsistent. For example, merging customer transaction records with a loyalty program database may result in missing loyalty IDs if some customers never enrolled.

Once you've identified missing data in your dataset, the next step is deciding how to deal with it. Let's explore some popular techniques to deal with missing values.

Techniques to Deal with Missing Values

Table 6-1 describes the most common techniques for managing missing values in data analysis workflows, when they can be used, and some practical examples of usage. The

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

correct approach depends on the nature of your data, the amount and pattern of missingness, and the type of analysis you plan to perform.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 6-1. The most common techniques for managing missing values

Technique	Description	Best for	Example
Deletion	Remove rows or columns. Row deletion is removing the (few) rows containing missing values. Column deletion is removing a column with a large portion of missing values.	There are a few missing rows (row deletion) or a significant portion of missing values in a column (column deletion).	If a survey response is missing for one participant, remove that row. If a column has too many missing values, consider removing the column.
Default or placeholder values	Fill missing cells with a standard placeholder, such as 0, N/A, or Unknown.	Non-critical fields, categorical columns, or when signaling that data was missing is useful.	Use “Unknown” for a missing gender entry or 0 for number of insurance claims.
Forward or backward fill	Forward fill uses the last known value to fill missing values below it. Backward fill propagates the next known value upwards.	Time-series data where the last observed state is assumed to continue temporarily.	If a stock price is missing for a day, you might assume it stayed the same as the previous day.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Technique	Description	Best for	Example
Interpolation	Estimates missing values based on nearby data points. This involves calculating a value that fits smoothly within the existing trend for numerical or time-series data.	Linear interpolation assumes a straight line between two known values.	Estimate a missing temperature reading by averaging the values before and after the missing one.
Statistical imputation	Replace missing values with statistical measures based on the column. Mean imputation calculates the average of existing values. Median imputation calculates the median. Mode imputation calculates the most frequent value.	Suitable for numerical columns when data is missing at random.	Use the mean income to fill in missing income values.
Conditional imputation	Infer missing values based on relationships with other fields.	Best for structured datasets where values can be inferred from other known columns.	Fill in the missing Branch Location based on the Salesperson.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Technique	Description	Best for	Example
Predictive modeling	Use machine learning models to predict missing values based on patterns in the rest of the data.	Complex datasets with enough data to train a reliable model.	Predict missing house prices using features like square footage, location, and number of rooms.

These strategies enable you to handle missing data; each has its strengths and appropriate contexts. You can combine different techniques for different dataset parts to achieve the most accurate and robust results.

Before moving to a practical exercise, let's see how Copilot deals with all these strategies. Table 6-2 shows the suggested prompt for Copilot for each technique and the possible Copilot answers. In fact, Copilot is not a deterministic tool; thus, its provided answer for each prompt may vary.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 6-2. Suggested prompts and possible Copilot answers for each technique to address missing values

Technique	Suggested prompt	Possible Copilot answers
Deletion	`Remove rows with missing values from the table. If a column has more than 50% missing values, drop it entirely.`	The steps to remove missing values The formula to remove missing values A button to directly remove rows with missing values and a button to remove columns with more than 50% missing values
Default or placeholder values	`Fill in the missing values in the X column with the value 'Unknown.'`	The formula to fill in the missing values Occasionally, Copilot adds a button to fill in the missing values directly
Forward or backward fill	`Use forward fill to propagate the last valid observation forward in column X. If forward fill fails, use backward fill.`	A button suggesting starting advanced analysis with Python The steps to apply this strategy with the formula
Interpolation	`Use linear interpolation to fill in missing values in the Value column.`	A button suggesting starting advanced analysis with Python The formula to apply interpolation

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Technique	Suggested prompt	Possible Copilot answers
Statistical imputation	`Impute missing values in the X column using the median of the non-missing values.`	A button suggesting starting advanced analysis with Python The formula to apply statistical imputation Occasionally, Copilot adds a button to directly apply imputation to the dataset
Conditional imputation	`Fill in missing values in the X column based on the <condition> in column Y.`	A button to directly apply conditional imputation to the table A statement saying that Copilot can't do that The formula to apply conditional imputation
Predictive modeling	`Train a classification model using available data to predict and fill missing values in the X column.`	A button suggesting starting advanced analysis with Python

Looking at the table, you will notice that in most cases, Copilot provides the formula and the steps to apply the required technique. Occasionally, it can apply the proposed technique directly to the data. In other cases, Copilot proposes starting advanced analysis. Click this button only if you have Python connected to Excel. Don't worry if, for now, you can't apply this strategy. You'll see this aspect later in the book.

Now that you've seen how Copilot's responses can vary, let's move on to a practical example. Don't worry if your results differ from those described in the example because Copilot doesn't generate answers deterministically, as mentioned earlier. What I'm offering is just one possible strategy for using Copilot. Refer to Table 6-2 to explore and compare alternative approaches. Ultimately, rely on your intuition to follow Copilot's suggestions to guide you toward a solution.

A Practical Example

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

When using Copilot to deal with missing values, the first step is to ask it to highlight them. Consider again the product sales dataset in the book's GitHub repository at *06/product_sales.xlsx*. Open the Copilot chat box by clicking the corresponding button in the Home ribbon. Write the following prompt for Copilot:

```
Highlight missing values
```

In most cases, Copilot should provide an answer similar to the one shown in Figure 6-6.

Figure 6-6. The answer provided by Copilot after asking it to highlight missing values

To apply this rule, click the Apply button. Figure 6-7 shows the result after applying the rule. The dataset contains two missing values in the `Branch Location` column.

Figure 6-7. The dataset with the highlighted missing value rows

Before asking Copilot to apply a specific strategy to deal with missing values, give it the following prompt:

```
Suggest actions for missing values.
```

This prompt is helpful to understand if Copilot already knows which strategy should be applied. In my case, Copilot provided the following answer:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
For the missing values in "Table1", here are some actions you can consider:
  1. Fill with default values: You can fill the missing "Branch Location" with a
     default value like "Unknown" or "Not specified".
     =IF([@[Branch Location]]="","Unknown",[@[Branch Location]])
  2. Use previous values: If the missing value is part of a series, you can fill
     it with the previous value in the column.
     =IF([@[Branch Location]]="",H9,[@[Branch Location]])
  3. Data validation: Ensure that future entries in the "Branch Location" column
     are mandatory to avoid missing values.
  4. Highlight for review: Keep the highlighted cells for manual review and
     correction.
These actions can help maintain the integrity and completeness of your data.
```

Here Copilot suggests strategies for default or placeholder values (1) and backward fill (2). In both cases, Copilot provides the formulas to apply. Copilot also proposes performing data validation (3) and highlighting cells with missing values (4). In our specific scenario, the suggested strategies (1) and (2) are not optimal because we can derive the Branch Location from the Salesperson (conditional imputation). Write the following prompt for Copilot:

```
Based on the Salesperson column, fill in the missing values in the Branch
Location column.
```

In this example, Copilot answers by proposing a formula and a button to insert the result directly into the table (Figure 6-8).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 6-8. The answer provided by Copilot when asked to apply conditional imputation

Next, click "Insert column." Figure 6-9 shows the resulting table, with the new column `Branch Location2` having only the first missing value (row 4) filled. The second missing value (row 6) is not filled.

Figure 6-9. The result of the "Insert column" button provided by Copilot to fill in missing values

The missing value in row 6 is replaced with a zero (0), although the filled value should be Boston (see row 10 for comparison). The problem here is that Copilot applies a formula that searches for the first occurrence in the `Branch Location` column corresponding to the Salesperson. The first occurrence in `Branch Location` for

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Salesperson David Green is an empty value, corresponding to zero (0). To solve the problem, ask Copilot to modify the formula:

```
Modify the following formula to search for the first occurrence of a non-empty
Branch Location corresponding to the Salesperson:
=IF([@[Branch Location]]="", XLOOKUP([@Salesperson], [Salesperson],
[Branch Location]), [@[Branch Location]])
```

In my case, Copilot provides the following answer:

```
=IF([@[Branch Location]]="", XLOOKUP([@Salesperson], FILTER([Salesperson],
[Branch Location]<>""), FILTER([Branch Location],
[Branch Location]<>""), ""), [@[Branch Location]])
```

Replace the original values in the column `Branch Location2` with the new formula. Figure 6-10 shows the result.

Figure 6-10. The resulting table after filling in the missing values

Now ask Copilot to explain the formula. In some cases, a "Show explanation" button appears. Simply click it to explain the formula. If you don't see the button, give Copilot the following prompt:

```
Explain this formula: =IF([@[Branch Location]]="", XLOOKUP([@Salesperson],
FILTER([Salesperson], [Branch Location]<>""), FILTER([Branch Location],
[Branch Location]<>""), ""), [@[Branch Location]]).
```

In my case, Copilot provided the following explanation:

```
Explanation: This formula fills in missing values in the "Branch Location" column
by finding the first non-empty branch location corresponding to the same
salesperson.
  • IF([@[Branch Location]]=""):
    Checks if the current row's "Branch Location" is empty.
    • If true, it proceeds to find the first non-empty branch location for the
      salesperson.
    • If false, it keeps the current "Branch Location" value.
  • XLOOKUP([@Salesperson], FILTER([Salesperson], [Branch Location]<>""),
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
FILTER([Branch Location], [Branch Location]<>""), ""):
• FILTER([Salesperson], [Branch Location]<>""): Filters the
  "Salesperson" column to include only rows where "Branch Location" is
  not empty.
• FILTER([Branch Location], [Branch Location]<>""): Filters the
  "Branch Location" column to include only non-empty values.
• XLOOKUP: Searches for the current row's salesperson ([@Salesperson])
  in the filtered "Salesperson" list and returns the corresponding value
  from the filtered "Branch Location" list. If no match is found, it
  returns an empty string ("").
• [@[Branch Location]]: If the "Branch Location" is not empty, it retains the
  current value.
```

TIP

Just for curiosity's sake, I also asked OpenAI's ChatGPT to explain the previous formula, using the same prompt, and I got a pretty similar answer to that provided by Copilot.

To confirm the theory that Copilot's response to the same prompt is not deterministic, I gave it the same prompt as input for the second time and got a different result. The starting prompt was this:

```
Based on the Salesperson column, fill in the missing values in the Branch
Location column.
```

Copilot provided the answer in Figure 6-11 in my second test.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 6-11. The answer provided by Copilot to deal with missing values in the second test

In practice, Copilot suggested applying the following formula:

```
=IF([@[Branch Location]]="",
    IF([@Salesperson]="Alice Smith", "New York",
    IF([@Salesperson]="Bob Jones", "Los Angeles",
    IF([@Salesperson]="Carol White", "Chicago",
    IF([@Salesperson]="David Green", "Boston",
    ""))))),
    [@[Branch Location]]
)
```

This formula includes all the possible values for `Salesperson`, preventing it from being generalizable. For example, if you add a new Salesperson to the table, the formula must be updated manually.

Exercise 1

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Consider the patient dataset shown in Figure 6-12, which is available in the book's GitHub repository at *06/patients.xlsx*.

Figure 6-12. The patient dataset

Using Copilot, perform the following tasks:

1. Highlight missing values.
2. Ask Copilot to suggest strategies for missing values in the Diagnosis and Blood Pressure columns.
3. Apply a strategy for filling in missing values in the Diagnosis and Blood Pressure columns.

A set of possible prompts for Copilot is available at *06/patients_prompts.txt*.

Data Formatting

When I buy a new product, like a phone or an appliance, I almost never read the instruction manual. And honestly, that's a bit of a problem. Caught up in the excitement, I usually rip open the packaging and start using it right away. But inevitably, something doesn't work as expected. That's when I go back, manual in hand, and skim through just the parts I need.

I have to admit, I approached Copilot the same way. I opened the chat box and started typing whatever came to mind. The results? Not exactly impressive. It was only after I hit a wall that I decided to dig deeper, by learning about prompt engineering, exploring online resources, and watching tutorials. And the difference? Huge.

So here's my advice: go ahead and experiment with your own prompts. That's part of the fun!

However, if you want to save time and avoid frustration, consider using proven prompts. They can give you a clearer idea of what Copilot is truly capable of and help you get better results, faster. In this section, you'll find a set of tested prompts specifically designed for data formatting tasks with Copilot.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Data formatting involves adjusting the structure, appearance, and consistency of data to ensure it's clean, readable, and ready for analysis. While it doesn't alter the underlying data itself, formatting plays a crucial role in how information is interpreted and utilized by both humans and software tools.

If your data isn't formatted correctly, you might run into problems like these:

Formula errors caused by text in numeric fields

> Excel formulas expect specific data types (numbers, dates, text). When the format is incorrect, calculations may silently fail or throw errors. For example, a number formatted as text may look fine but won't behave correctly in operations like SUM, AVERAGE, or mathematical comparisons.

Incorrect aggregations due to inconsistent labels

> When labels vary in capitalization, spacing, or spelling, Excel treats them as distinct entries. This can lead to misleading summaries in PivotTables or grouped calculations. For example, if a column contains "Q1," "q1," and " Q1 " (with a leading and trailing space), Excel will group these as separate categories in a PivotTable, thus fragmenting what should be a single group.

Common issues in data formatting include inconsistent text or number formats, dates stored as text, merged cells, and hidden characters, all of which can disrupt calculations, filtering, and readability. Table 6-3 shows the most common data formatting problems in Excel.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 6-3. Common data formatting problems in Excel

Data type	Formatting problem	Description
Textual	Inconsistent text formatting	Varying use of case, spacing, special characters, or line breaks
	Inconsistent use of symbols	Units or symbols embedded in text inconsistently (e.g., “50kg” vs. “50 kg”)
Non-textual	Number stored as text	Numeric-looking values stored as text (e.g., “1000”)
	Date and time format issues	Dates stored as text or mixed formats; locale inconsistencies
	Inconsistent number formatting	Differences in decimal places, currency symbols, or thousands separators
Both	Hidden characters or formatting	Tabs, line breaks, or conditional formatting hiding content
	Misuse of cell formatting for meaning	Using color, bold, etc. without logic or metadata backing
	Blank or missing values misrepresented	Use of “N/A,” “-,” or invisible characters instead of true blanks

Techniques to Deal with Data Formatting Issues

Table 6-4 describes the most common techniques for managing data formatting issues and some practical examples of usage.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 6-4. The most common techniques for managing data formatting issues

Technique	Description	Example
Standardize text	Making textual data consistent in structure, spacing, and capitalization	`Convert " john DOE " to "John Doe".`
Split text from symbols	Separating mixed data (e.g., values and units) into distinct components	`Split "50kg" into 50 and kg.`
Convert text to numbers	Changing string representations of numbers into numeric values	`Change '1000' to 1000 for calculations.`
Unify date formats	Ensuring all date entries follow a consistent format and are recognized as dates	`Convert "01/02/2023" to a standard date format.`
Make number formatting consistent	Applying uniform formatting to numerical values (e.g., decimals, currency)	`Format a column as currency with two decimals.`
Clear hidden characters	Removing invisible or nonprintable characters like tabs, line breaks, or spaces	`Clean data copied from web or PDF sources.`
Use consistent cell formatting for meaning	Using color, bold, or other formatting to convey information without logical rules or data backing	`Replace manual highlighting with a rule-based color scale (e.g., red for values < 50).`
Replace placeholders	Substituting placeholder entries like “N/A” or “-” with standardized alternatives or blanks	`Replace "N/A" with a blank or default label.`

The techniques described are not exhaustive, but they provide an idea of how to deal with data formatting problems.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Before moving on to a practical example, let's investigate how you can use Copilot to apply the described techniques. Table 6-5 presents a suggested prompt for each technique, along with the possible answers provided by Copilot.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 6-5. Suggested prompts and possible Copilot answers for each technique to address data formatting issues

Technique	Suggested prompt	Possible Copilot answers
Standardize text	`Standardize the text in column X: trim extra spaces and use proper capitalization.`	A formula using `=PROPER(TRIM(@X))` A column preview with cleaned-up names, with a button to apply formatting. A formula using `=TEXTJOIN(" ",TRUE,PROPER(TRIM(Table[@X])))`
Split text from symbols	`Split column X into two separate columns: numeric values and units (e.g., "50kg" into "50" and "kg").`	A preview table with the results and an “Insert columns” button Two formulas, one for each column, such as: `=IFERROR(LOOKUP(9.9E+307,--TEXTSPLIT([@X],{"kg","g","mg"})),0)` `=IFERROR(TEXTAFTER([@X],TEXT(LOOKUP(9.9E+307,--TEXTSPLIT([@X],{"kg","g","mg"})),"0")),[@X])`
Convert text to numbers	`Convert all text-formatted numbers in column X to actual numbers.`	A button to apply the following changes: number format on column X; Autofit column(s)
Unify date formats	`Convert all dates in column X to the format YYYY/MM/DD.`	A button to apply the following changes: date/time on column X; Autofit column(s)
Make number formatting consistent	`Format all numbers in column X to show two decimal places and use comma separators.`	A button to apply “Number” format with specific decimal and thousands separators

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Technique	Suggested prompt	Possible Copilot answers
Clear hidden characters	`Clean hidden characters (e.g., line breaks, tabs) from column X.`	A formula using `=TRIM(CLEAN([@X]))` Occasionally, Copilot incorrectly proposes a button to apply the following change: Clear the content of some cells
Use consistent cell formatting for meaning	`Create a conditional formatting rule for column X: highlight values based on thresholds: <10 red, between 11 and 100 yellow, > 101 green.`	A button to apply the rule
Replace placeholders	`Replace placeholder values like 'N/A' or “-” in column X with blank cells (or a default value).`	A find and replace operation A formula using `=IF(OR([@X]="N/A",[@X]="-"),"",[@X])` Occasionally, Copilot incorrectly proposes a button to apply the following change: Clear the content of some cells

Alternatively, Copilot provides another feature to manage data formatting issues, namely the Clean Data button in the Data tab of the Ribbon (Figure 6-13).

Figure 6-13. The Clean Data button in the Data tab of the Ribbon

Clicking this button enables you to solve some data formatting issues, including removing extra spaces, making text in cells consistent, and formatting number columns consistently.

At this point, you are ready to practice with an example.

A Practical Example

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Consider the sample movie dataset in the book's GitHub repository, located at *06/movies.xlsx* and illustrated in Figure 6-14.

Figure 6-14. A preview of the sample movie dataset

The dataset contains information about individuals' movie preferences and their feedback. It includes the following columns:

`Name`

: The name of the individual

`Age`

: The individual's age

`Favorite Genre`

: The preferred movie genre of the individual

`Rating`

: The individual's rating of the movie

`Feedback`

: Textual comments about the movie

`Date Watched`

: The date when the movie was watched

The dataset contains several inconsistencies, such as inconsistent capitalization, text-based ratings, and missing values. Let's use Copilot to address these problems. Open

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

the Copilot chat box by clicking the corresponding button in the Home tab of the Ribbon. Write the following prompt for Copilot:

```
Standardize the text in the column Name: trim extra spaces and use proper
capitalization.
```

In my case, Copilot answered with the output shown in Figure 6-15.

Figure 6-15. The answer provided by Copilot after asking it to standardize text

Click “Insert column” or manually apply the formula. This will convert entries like " JOHN doe " to “John Doe” and remove extra spaces or capitalization inconsistencies. Figure 6-16 shows the resulting column with the standardized names.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 6-16. The movie dataset with a new column for the standardized names

Next, we address another common formatting issue: numbers stored as text. Write the following prompt for Copilot:

```
Convert all text-formatted numbers in the Age column to actual numeric values.
```

In my case, Copilot provided the answer shown in Figure 6-17.

Figure 6-17. The answer provided by Copilot to convert all text-formatted numbers to actual numeric values

Apply the formula or follow the instructions to convert values. Figure 6-18 shows the cleaned numeric column. If you have a different experience with Copilot, don’t worry,

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

because you can obtain the same result by following alternative steps.

Note that Copilot has replaced the missing value for Jane Doe with 0 (zero). This could be a problem, so before solving data formatting issues, I suggest that you address missing values.

Now, fix the date format inconsistency in the `Date Watched` column. Use this prompt:

```
Convert all dates in the column Date Watched to the format YYYY-MM-DD.
```

In my case, Copilot provided the answer shown in Figure 6-19.

Figure 6-18. The movie dataset with a new column for the converted age

Figure 6-19. The answer provided by Copilot to change the date format

After applying the suggested changes, you should obtain the result shown in Figure 6-20.

Figure 6-20. The movie dataset with a new column for the converted Date atched format

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

As shown in the figure, Copilot correctly applied the formatting only to the cells that were already recognized as dates. However, the second row remains incorrectly formatted. To fix this problem, write the following prompt for Copilot:

```
Suggest how to remove uotation marks from the column Date Watched.
```

The solution proposed by Copilot in my case didn't solve the problem. Thus, I engaged in the conversation with Copilot that is shown in Figure 6-21.

Figure 6-21. The conversation with Copilot to solve the date formatting problem

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

When I prompted Copilot that its proposed formula didn't work, Copilot immediately proposed the following alternative formula:

```
=SUBSTITUTE(TEXT([@[Date Watched]],"yyyy-mm-dd"),"""","")
```

Figure 6-22 shows the result after applying the previous formula to a new column.

Figure 6-22. l is a new column with corrected dates

Looking at the date in row 8, you will notice that Copilot has transformed an empty date into 1900-01-00. You can ask Copilot to fix this, for example, by giving it the following prompt:

```
Modify the following formula so that empty rows are left empty:
=SUBSTITUTE(TEXT([@[Date Watched]],"yyyy-mm-dd"),"""","")
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

A T MATIC DETECTI F C MM ERR RS

When you open a workbook with formatting issues, Copilot can automatically detect them and suggest corrections through a notification (Figure 6-23).

Figure 6-23. hen you open a new workbook with formatting issues, Copilot may suggest cleaning your data

Click the “Show suggestions” button, and the Copilot chat box opens. You should see the two boxes shown in Figure 6-24.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 6-24. The suggestions provided by Copilot to clean the dataset

In my case, Copilot suggested correcting the inconsistent number format in column B and inconsistent text in column C. I clicked Apply in both cases. The result is shown in Figure 6-25.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 6-25. The movie dataset with columns and C cleaned

Exercise

The movie dataset shown in Figure 6-25 still contains some formatting problems in the `Favorite Genre, Rating`, and `Feedback` columns. Use Copilot to format the dataset properly.

Exercise

Consider again the patient dataset available in the book's GitHub repository at *06/patients.xlsx*.

Using Copilot, perform the following tasks:

1. After removing the quotation marks, format the Age column as numbers. If needed, align values to the right.
2. After removing quotation marks, format the Billing Amount column as currency.
3. Format the Admission Date column properly.

A set of possible prompts for Copilot is available at *06/patients_prompts_formatting.txt*.

Challenge Is your dataset now cleaned If not, identify what is still required to format your dataset properly and ask Copilot to clean it properly.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

KEEP THE FIRE ALIVE

Perhaps you began this journey filled with excitement. Or maybe with a bit of hesitation.

Maybe you opened Copilot hoping for instant answers, only to find confusion, silence, or unexpected results. And if that happened, let me say this clearly: that's okay. That is learning. Learning to work with data and with a tool like Copilot isn't about mastering everything at once. It's about patience. It's about persistence. It's about meeting resistance and choosing to try again.

I'm reminded of a winter day from my childhood, in a house tucked high in the mountains.

We heated our home with a wood-burning fireplace, and one afternoon my mother asked me to tend the fire. Proud and confident, I took the tools in hand…and did it all wrong. Logs moved aimlessly, smoke thickened, and the flame nearly died. My mother gently offered guidance, but I refused. I wanted to do it my way. In the end, she stepped in, fixed it, and I stood by, arms crossed, not because I had learned, but because I hadn't let myself learn.

It took me years to understand what really failed me that day: not my technique, but my pride.

To learn, we must first accept that someone might know more than we do. And that's not weakness. That's growth.

If you've struggled with Copilot, if it hasn't worked the way you hoped, don't give up.

Let go of the need to be perfect. Be curious. Be humble. Ask questions. Search for answers.

There's no shame in learning from others. On the contrary, it's the fastest path forward.

Because every prompt you write, every small frustration, every tweak and retry, they're sparks.

Sparks that will catch. Sparks that will become momentum. That's how we move from confusion to clarity, from doubt to discovery. So don't stop now. Keep going, one step, one prompt, one breakthrough at a time.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Summary

In this chapter, you learned how to use Copilot within Excel to prepare data through three key tasks: removing duplicates, dealing with missing values, and formatting:

Copilot helps clean your data by identifying and highlighting duplicate entries. While it can't remove duplicates automatically, it guides you through Excel's built-in tools for doing so, ensuring data integrity before analysis.

Copilot helps you handle missing values by suggesting strategies such as deletion, default replacements, and imputation. Depending on the context, Copilot offers formulas, interactive buttons, or Python integration to effectively fill in gaps.

You can use Copilot to fix data formatting issues such as inconsistent text, dates stored as text, or numbers formatted incorrectly. It recommends formulas and applies standard formatting to make your data consistent and analysis-ready.

In the next chapter, you'll learn how to transform text with Copilot.

Activity

Complete the following text using the keywords indicated below. The solution is in Appendix D.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

COPILOT

DATA FORMATTING

DUPLICATES

INSIGHTS

MISSING VALUES

HIGHLIGHTING RULES

PREPARATION

PROMPT

In this chapter, you learned how important proper data is before any analysis.

To support this process, you explored how to structure an effective when interacting with inside Excel.

Three core operations were covered: identifying and removing , addressing , and applying consistent .

Each of these steps is essential to eliminate noise, correct errors, and ensure that your data can generate reliable .

For instance, giving Copilot clear and specific instructions allows it to respond more effectively to your needs. Meanwhile, tools such as and customized formulas support the reliability and consistency of your dataset.

Keep in mind: preparing data isn't just a technical step; it's what allows patterns to surface, errors to be minimized, and meaningful conclusions to be drawn with confidence.

Chapter 7. Text Transformation and Manipulation

A few days ago, I found myself unexpectedly flipping the script. It was a quiet afternoon. I was out on a walk with my family, my kids pedaling along happily on their bikes. Everything seemed fine until suddenly, it wasn't. Without warning, they grew tired and bored. The mood shifted. The whining began. That's when I made a decision. I turned to Giulia and asked her, much to her surprise, if I could borrow her bike. Now, let me clarify: this was a child's bike, far from my-age-appropriate. And I hadn't ridden in over a decade, not since my last bicycle was stolen. But something told me this moment needed a change. As soon as I started pedaling, something stirred. The sensation of riding, the rhythm, the balance, it all came flooding back. And with it came a forgotten joy: seeing the world from the saddle. Trees stretched above, the sky opened wide in sharp blue, and suddenly, the mundane became wonderful. I was smiling. Laughing. Fully present. And that joy? It was contagious. My kids, moments earlier bored and fussy, lit up. They started chasing after me, giggling, eager to join the ride again. What had been a dull afternoon turned into a shared adventure.

This, believe it or not, is what working with Copilot can feel like. Copilot has the power to flip the script on how you approach tedious tasks. It can help you see a boring, repetitive job with new eyes. In addition, it can make those very tasks interesting, even enjoyable. Just as riding that tiny bike reignited my curiosity and joy, using Copilot to write and manipulate text formulas or troubleshoot errors can shift your experience from passive to engaged. What once felt routine becomes dynamic. You might even find yourself looking forward to solving those logic puzzles you used to avoid.

In this chapter, we'll explore exactly how AI can help you do that. Many of the ChatGPT skills you learned in earlier chapters also apply to Copilot. We'll cover:

- Explaining text formulas
- Generating text formulas
- Handling errors
- Applying logic to text

Let's start with the first point: explaining text formulas.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Explaining Text Formulas

Interpreting formulas can be challenging, especially when they aren't well documented. Even with our own work, it's easy to lose track of what a formula was intended to do after some time. The challenge becomes even greater when working with spreadsheets created by others, where the logic behind the formulas may be completely unfamiliar.

This is where Copilot becomes especially useful. It serves as a real-time interpreter, helping you understand even the most complex formulas. Copilot can explain the purpose of a formula in its specific context, break down each step of a transformation, and rephrase technical logic in clear, shareable English.

As discussed in the previous chapter, choose the Copilot modality based on your intent: use Agent Mode to modify the worksheet, Analysis Mode to compute metrics without changing the data, and chat for general questions. In this book, however, the focus is not on specific feature names, since the interface evolves, but on the underlying principles, so you can adapt confidently to future versions of Copilot.

You can interact with Copilot using natural language. Here's a helpful way to structure your prompt:

Main point

> Decide what you want Copilot to do (for example, explain something).

Context

> Tell Copilot where the formula is located (a specific cell or column) or give it the formula itself.

For example, you can write the following prompt for Copilot:

```
Explain the formula in column A
```

This approach enables you to quickly gain clarity on spreadsheet logic, without needing to dig through documentation or reverse-engineer someone else's work.

To better understand how this prompt structure works, consider the Excel file available in the GitHub repository for the book at *07/books.xlsx*. Figure 7-1 shows the first rows of the table contained in the first worksheet of the dataset.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 7-1. A portion of the books.xlsx dataset

The table contains the following columns:

Title

The full name of the book; often includes subtitles to clarify the content or focus

Author

The individual or group credited with writing the book; listed by full name

Year

The publication year of the book; indicates when the book was published

Summary

A brief description of the book's content; may include the scope, target audience, or central themes

Topics

Key themes or subjects covered in the book

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

The second worksheet in the workbook, named Topics, contains a list of topics, as shown in Figure 7-2.

Figure 7-2. The Topics worksheet contains a single-column table named Topics

Select the Books worksheet and double-click cell E2. The following formula should appear:

```
=LET(
  summary, D2,
  topics, Topics!A$2:A$6,
  matches, IF(ISNUMBER(SEARCH(topics, summary)), topics, ""),
  result, TEXTJOIN(", ", TRUE, matches),
  IF(result = "", "Other", result)
)
```

This formula is quite complex, so you can use Copilot to explain it. Open the Copilot chat box by clicking the related icon on the Home tab of the Ribbon. Write the following prompt for Copilot:

```
Explain the formula in the column Topics of the worksheet Books.
```

In my case, Copilot answered with the following detailed response:

```
The formula in the "Topics" column identifies relevant topics mentioned in the
"Summary" column and combines them into a single string. If no topics are
found, it defaults to "Other." Here's the formula:
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Explanation:

- LET(summary, D2, topics, Topics!A$2:A$6, ...): Assigns the "Summary" text of the current row to summary and the list of predefined topics (range Topics!A$2:A$6) to topics.
- SEARCH(topics, summary): Checks if each topic in the topics list appears in the summary. Returns a number if found, or an error if not.
- ISNUMBER(SEARCH(...)): Converts the result of SEARCH into TRUE for matches and FALSE for non-matches.
- IF(ISNUMBER(...), topics, ""): Keeps the topic if it matches, otherwise returns an empty string.
- TEXTJOIN(", ", TRUE, matches): Combines all matching topics into a single string, separated by commas, ignoring empty strings.
- IF(result = "", "Other", result): If no topics are matched, returns "Other"; otherwise, returns the combined topics.

This formula dynamically extracts and lists relevant topics based on the content of the "Summary" column.

Generating Text Formulas

When you know what you want to do with text data but can't quite remember the exact Excel formula, Copilot can help you. Just describe your goal in plain English, and Copilot will generate the formula for you. To get accurate and relevant results, frame your prompt like a short, focused story. This gives Copilot enough context to return a useful response.

Here's a simple structure to follow when writing your prompt:

Context

What is the data about? What column or table are you working with?

Main point

What do you want to achieve or extract from the text?

Action (optional)

Any format or condition you want to request.

To understand how this structure works, consider the *books.xlsx* dataset again. Ask Copilot to write the formula to split the list of authors in the Author column into different columns, one for each author. You can organize the prompt using the context–main point–end structure. Give Copilot the following prompt:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
Consider the Author column. Write the formula to divide the authors into as many
columns as there are authors (use a comma or the string ' and ' as a separator).
Add each author to a new column: First Author, Second Author, Third Author,
and so on.
```

Let's break down the prompt into the story elements:

Context

> Consider the Author column.

Main point

> Write the formula to divide the authors into as many columns as there are authors (use a comma or the string ' and ' as a separator).

Action

> Add each author to a new column: First Author, Second Author, Third Author, and so on.

In my case, Copilot proposed two alternatives: manually adding formulas to the table, or getting Copilot to insert them for you. The first alternative includes all the formulas to extract the first, second, and third authors:

> First author: `=TRIM(INDEX(TEXTSPLIT(SUBSTITUTE([@Author]," and ",","),","),1,1))`
>
> Second author:
> `=TRIM(IFERROR(INDEX(TEXTSPLIT(SUBSTITUTE([@Author]," and ",","),","),1,2),""))`
>
> Third author:
> `=TRIM(IFERROR(INDEX(TEXTSPLIT(SUBSTITUTE([@Author]," and ",","),","),1,3),""))`

You can paste these three formulas into three different columns in row 2, and manually change the column names to First Author, Second Author, Third Author. This operation extracts the authors for only row 2. To extend the formulas to the other rows, select the three cells containing the formulas (e.g., E2, F2, and G2 if that's where you pasted them), then drag the Fill Handle (a small square at the bottom-right corner of the

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

selection) downward to fill the formulas into the rows below. Figure 7-3 shows the final result.

Figure 7-3. The final table, after adding the three columns related to the books' authors

The table now has three new columns, one for each author.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

COPILOT, EXPLAIN YOURSELF

Ask Copilot to explain the first formula, using the following prompt:

```
Explain the following formula:
=TRIM(INDEX(TEXTSPLIT(SUBSTITUTE([@Author]," and ",", "),",",), 1, 1))
```

In my case, Copilot answered as follows:

> This formula extracts the first author's name from the "Author" column, even if the names are separated by commas or the string " and ". Here's how it works step-by-step:
>
> - SUBSTITUTE([@Author]," and ",", "): Replaces all occurrences of " and " in the "Author" field with ", ". This ensures that both " and " and commas are treated as delimiters.
> - TEXTSPLIT(..., ","): Splits the modified text into an array of names using a comma (,) as the delimiter.
> - INDEX(...,1,1): Retrieves the first item (row 1, column 1) from the resulting array, which corresponds to the first author's name.
> - TRIM(...): Removes any leading or trailing spaces from the extracted name to ensure a clean result.
>
> In summary, this formula extracts and cleans the first author's name from the "Author" column, regardless of whether the names are separated by commas or " and ".

You can repeat the prompt to ask Copilot to explain the other formulas.

The second alternative provided by Copilot is a prefilled table shown directly in the chat box with an “Insert columns” button, as shown in Figure 7-4.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 7-4. The option to add the columns directly into the table through the "Insert columns" button

Click the "Insert columns" button to obtain the result shown back in Figure 7-3.

So far, you have seen how to split text in one cell into multiple cells. However, Excel provides many other text transformations. Some of them are described in Table 7-1.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 7-1. The most common categories of text transformation

Transformation type	Example prompt for Copilot	Formula example

Split operation	Split the Author column into two columns, Name and Surname.	Extract name:
		`=TEXTBEFORE([@Author]," ")`
		Extract surname:
		`=TEXTAFTER([@Author]," ")`

Extraction	Extract the first 20 characters in the Summary column.	`=LEFT([@Summary],20)`
Substring and pattern	Get the first sentence contained in the Summary column.	`=IF(ISNUMBER(FIND(".",[@Summary])),TEXTBEFORE([@Summary],".",1,,,[@Summary])&".",[@Summary])`
Concatenation and formatting	Create an ID using author and year with no spaces, all lowercase.	`=LOWER(SUBSTITUTE([@Author]&[@Year]," ",""))`
Cleanup and normalization	Lowercase the first letter of each word in the Topics column.	`=TEXTJOIN(", ",TRUE,LOWER(TRIM(TEXTSPLIT([@Topics],","))))`

Exercise 1

Consider the *tech_sales.xlsx* workbook available in the GitHub repository for the book at *07/tech_sales.xlsx*. Figure 7-5 shows the table contained in the workbook.

Figure 7-5. The table contained in the tech_sales.xlsx workbook

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

The table contains the following columns:

Customer Name

Full name of the customer, with possible variations in capitalization and spacing

Email

Customer's email address; some entries may be missing or incomplete

Product

Name of the product purchased by the customer, such as Laptop Pro or Wireless Mouse

Category

Broad classification of the product, such as Computing, Accessories, or Displays

Job Title

Professional role or title of the customer, such as developer, CTO, or intern

Date

The date of purchase

Write prompts for Copilot to do the following:

1. Split the Customer Name column into two columns, Name and Surname. Capitalize the first letter in the name and surname, and keep the other letters lowercase. For surnames like McGuillan, McDonald, or O'Connor, the internal capital letter must also be preserved (e.g., McGuillan, not Mcguillan).
2. Create an ID using the customer's name and date of purchase.

You can find the solution to the exercise in the GitHub repository for the book at *07/exercise1_solution.txt*.

So far, you've seen how to use Copilot to explain and generate text formulas. But in real-world spreadsheets, data is rarely perfect. The formulas that Copilot generates by

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

default often assume clean, consistent input, which is an ideal condition that doesn't always reflect reality.

Think of it like tailoring doll clothes. When I was a child, I liked making outfits for my dolls. I often followed patterns drawn on paper, but those patterns were too generic. When I tried them on a specific doll, they rarely fit exactly right. Every doll was different, so I'd have to take out my needle and thread to adjust the clothes until they fit correctly.

Formulas work the same way. A general formula might look fine on paper, but when applied to real, messy data, it often needs adjustments. That's where Copilot can help: you can guide it to create formulas that account for errors, inconsistencies, and exceptions so the result fits your data just right.

In the next section, you'll learn how to prompt Copilot to handle errors and edge cases.

Handling Errors

Real-world data often contains inconsistencies, such as missing values, incomplete entries, or incorrect formats. This can cause formulas to break or return errors like #VALUE! or #N/A. Copilot can help prevent and handle these cases using error-handling functions like IFERROR() and IF(). For error handling, you can use two different approaches:

Describe the issue

> Explain the specific problem in your dataset to Copilot and ask it to generate a formula that can handle that issue.

Let Copilot detect problems

> Allow Copilot to analyze your data and automatically generate formulas that are robust against the issues it identifies.

Let's analyze each approach separately, starting with a description of the issue.

Describing the Issue

When you describe the issue in your data, you're starting from a known problem. You've already identified what's wrong, and now you must describe it in natural

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

language to get Copilot to understand and handle it. Copilot will analyze the context and suggest ways to make your formulas more robust and error-resistant.

To get accurate results, structure your prompt using the story structure, with the following three key elements:

Context

> Describe where the issue occurs. Example: `Column B contains cells with missing values.`

Main point

> Define the goal of your formula. Example: `Write a formula to fill in missing values with a default.`

Action

> Explain how to handle the errors. Example: `Identify missing values by looking for cells marked as NaN or empty.`

To understand how this prompt works, consider the dataset *books_with_errors.xlsx* available in the GitHub repository for the book at *07/books_with_errors.xlsx* and shown in Figure 7-6.

The dataset contains several errors. For now, consider the error in the `Author` column in which the space between the name and surname is missing for row 2. Open the workbook in Excel, access the Copilot chat box, and write the following prompt:

```
Column Author contains errors because the space between name and surname is
sometimes missing. Write the formula to divide the Author into two columns:
Name and Surname. Handle cases where the formatting is incorrect by using
capital letters to split Name and Surname.
```

Figure 7-6. The table contained in the books_with_errors.xlsx workbook

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Let's break down the prompt into the key elements:

Context

> Column Author contains errors because the space between name and surname is sometimes missing.

Main point

> Write the formula to divide the Author into two columns: Name and Surname.

Action

> Handle cases where the formatting is incorrect by using capital letters to split Name and Surname.

In my case, Copilot provided the following output:

> Name: =IFERROR(TEXTBEFORE([@Author],"
> "),LEFT([@Author],MIN(IFERROR(FIND(CHAR(ROW(INDIRECT("66:90")))
>),[@Author],2),LEN([@Author])+1))-1))
>
> Surname: =IFERROR(TEXTAFTER([@Author],"
> "),MID([@Author],MIN(IFERROR(FIND(CHAR(ROW(INDIRECT("66:90"))))
> ,[@Author],2),LEN([@Author])+1)),LEN([@Author])))

Additionally, Copilot built the table and provided an "Insert columns" button to add the two columns directly to the table (Figure 7-7).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 7-7. Copilot's option to add the columns directly into the table through the "Insert columns" button

Ask Copilot to explain the first formula, using the following prompt:

```
Explain the following formula: =IFERROR(TEXTBEFORE([@Author]," "),LEFT([@Author],
MIN(IFERROR(FIND(CHAR(ROW(INDIRECT("66:90"))),[@Author],2),LEN([@Author])+1))-1))
```

In my case, Copilot answered as follows:

```
This formula extracts the first name or given name from the "Author" column, even
if the name is formatted incorrectly (e.g., missing a space between the first and
last name). Here's how it works:
Explanation:
  1. TEXTBEFORE([@Author]," "):
     • Attempts to extract the text before the first space in the "Author" value.
       This works for names formatted as "First Last".
  2. IFERROR(..., ...):
     • If there is no space in the "Author" value (e.g., "BrentDykes"), the
       formula moves to the second part of the IFERROR function.
  3. LEFT([@Author],MIN(...)-1):
     • Extracts characters from the start of the "Author" value up to the position
       of the first uppercase letter (other than the first character), assuming
       the uppercase letter marks the start of the last name
  4. FIND(CHAR(ROW(INDIRECT("66:90"))),[@Author],2):
     • Searches for the position of any uppercase letter (ASCII codes 66 to 90,
       corresponding to "B" through "Z") in the "Author" value, starting from
       the second character. This handles cases where there is no space but
       the last name starts with an uppercase letter.
  5. MIN(IFERROR(...,LEN([@Author])+1)):
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

- Finds the earliest position of an uppercase letter. If no uppercase letter is found, it defaults to one more than the length of the "Author" value, ensuring the entire name is treated as the first name.

6. IFERROR(..., ...):
   - Ensures the formula does not break if no uppercase letter is found.

Summary:

- If the "Author" value contains a space, the formula extracts the text before the space.
- If there is no space, it extracts characters up to the first uppercase letter after the first character, assuming this marks the start of the last name.

You can repeat the prompt to ask Copilot to explain the other formula.

The previous example illustrated a specific type of error and how to handle it. More generally, there are several categories of common errors you might encounter. Table 7-2 summarizes the most frequent error types, their potential consequences, and examples of Copilot prompts you can use to address them.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 7-2. Frequent data error types in Excel

Error type	What could go wrong	Example of a prompt to handle the error
Missing expected text structure	Text separators like spaces or commas are missing, causing formulas like FIND() or TEXTSPLIT() to break.	Column Author contains errors because the space between name and surname is sometimes missing. Write the formula to divide the Author into two columns: Name and Surname. Handle cases where the formatting is incorrect by using capital letters to split Name and Surname.
Blank or missing input values	Fields are left blank, causing formulas like LEN() or PROPER() to return blanks or errors.	Column Topics contains errors because some cells are empty. Write the formula to extract the first topic from non-empty cells. Handle blank cells by returning a custom value like None.
Inconsistent formatting	Casing, spacing, or indentation varies, making uniform text processing difficult.	Column Title contains errors because the casing is inconsistent (e.g., all caps or mixed case). Write the formula to standardize the casing. Handle different formats and return the result in title case.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Error type	What could go wrong	Example of a prompt to handle the error
Special characters (e.g., apostrophes)	Irregular apostrophes, quotation marks, or encoding issues cause parsing problems.	`Column Year contains errors due to special characters and encoding issues (e.g., curly quotes, encoded apostrophes). Write the formula to clean these characters. Replace corrupted characters with standard equivalents.`
Unexpected data types (e.g., numbers)	Text columns unexpectedly contain numbers, which can confuse string functions.	`Column Status contains errors because some cells have numbers instead of text. Write a formula that checks if the value is a number and converts it as follows: if the value is 0, return To read; if the value is 1, return Read. Handle other cases by returning Unknown.`
Dependent values are missing or incomplete	When combining fields like `Title` and `Year`, a missing component results in broken outputs.	`Column Author contains the full name (first name and surname), but in some cases, the space between them is missing. Write a formula to extract the first name by using a space if available, or by detecting the character before the first capital letter after the first character if the names are merged. Then build an ID by combining the extracted first name, a dash (-), and the current row number.`

Note how the prompts listed in Table 7-2 follow a clear structure: context—main point—end.

Exercise 2

Revisit the *books_with_errors.xlsx* dataset. Open it in Excel, launch Copilot, and test whether the prompts provided in Table 7-2 actually address the corresponding errors.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Challenge: The output generated by the prompt for the inconsistent formatting error may not be entirely correct. Can you identify what may be wrong and suggest a better way to handle it?

Letting Copilot Detect Problems

You don't always know exactly what's wrong with your data. In that case, you can ask Copilot to help identify potential issues for you. Start by specifying which column contains errors or unexpected values. Then ask Copilot to analyze the data and suggest a strategy for handling the problem. Once a plan is in place, you can ask it to generate the formula needed to implement the solution.

As with the previous cases, you can structure your prompt as a short story, with the following key elements:

Context

Describe where the issue occurs.

Main point

Ask Copilot to identify errors and elaborate a strategy to handle them.

Action

Ask Copilot to perform a specific task, such as generating a formula.

To understand how this prompt structure works, consider the *books_with_errors.xlsx* workbook again. Write the following prompt for Copilot:

```
Column Author contains errors. Identify the type of errors and elaborate on a
strategy to handle them. Write the formula to divide the Author into two
columns: Name and Surname.
```

Let's break down the prompt into the key elements:

Context

Column Author contains errors.

Main point

Identify the type of errors and elaborate on a strategy to handle them

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Action

Write the formula to divide the Author into two columns: Name and Surname.

In my case, Copilot provided the following response:

```
The "Author" column contains errors such as:
  1. Missing spaces: Some entries (e.g., "BrentDykes") have no space between the
     first name and surname.
  2. Inconsistent formatting: Some entries may have extra characters or unusual
     spacing.
Strategy to handle errors:
  1. Detect missing spaces: Use logic to identify the first uppercase letter
     after the first character to split names when spaces are missing.
  2. Standardize formatting: Ensure consistent splitting into "Name" and
     "Surname" columns using formulas.
Formulas to divide "Author" into "Name" and "Surname":
Name:
Extract the first name by using the space if available, or by detecting the first
uppercase letter after the first character.
=IFERROR(TEXTBEFORE([@Author]," "),LEFT([@Author],
MIN(FILTER(SEQUENCE(LEN([@Author])-1,1,2,1),ISNUMBER(FIND(MID([@Author],
SEQUENCE(LEN([@Author])-1,1,2,1),1),"ABCDEFGHIJKLMNOPQRSTUVWXYZ"))))-1))
Surname:
Extract the surname by using the text after the space if available, or by
detecting the first uppercase letter after the first character.
=IFERROR(TEXTAFTER([@Author]," "),MID([@Author],
MIN(FILTER(SEQUENCE(LEN([@Author])-1,1,2,1),ISNUMBER(FIND(MID([@Author],
SEQUENCE(LEN([@Author])-1,1,2,1),1),"ABCDEFGHIJKLMNOPQRSTUVWXYZ")))),
LEN([@Author])))
These formulas handle both cases (with and without spaces) and ensure consistent
splitting into "Name" and "Surname".
```

To apply the strategy proposed by Copilot, insert a new column directly to the right of your existing table in your worksheet (this should be the first column outside the table), so Excel can automatically expand the table when you add the formula. Click the second row of this new column (the first row is the header) and enter the following formula proposed by Copilot to extract the Name:

```
=IFERROR(TEXTBEFORE([@Author], " "),
    LEFT([@Author], MIN(FILTER(
        SEQUENCE(LEN([@Author])-1, 1, 2),
        ISNUMBER(FIND(MID([@Author], SEQUENCE(LEN([@Author])-1, 1, 2), 1),
        "ABCDEFGHIJKLMNOPQRSTUVWXYZ"))
    )) - 1)
)
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Press Enter. Excel will fill the entire column and extend the table to include it. Repeat the process in the next column to extract the Surname, using the following formula:

```
=IFERROR(TEXTAFTER([@Author]," "),
  MID([@Author],MIN(FILTER(
    SEQUENCE(LEN([@Author])-1,1,2),
ISNUMBER(FIND(MID([@Author],SEQUENCE(LEN([@Author])-1,1,2),1),
    "ABCDEFGHIJKLMNOPQRSTUVWXYZ")))),
  LEN([@Author])))
```

With these two formulas in place, your table will now include cleanly separated Name and Surname columns.

Applying Conditional Text Logic

Conditional text logic is used when you want to assign labels, tags, or outcomes based on the contents of a cell. With Copilot, you can describe the condition in plain English, and it will generate the correct formulas for you.

Consider the *books.xlsx* dataset again, available in the book's GitHub repository at *07/books.xlsx*. Write the following prompt for Copilot to label data based on the text in the Title column:

```
Consider the Title column. Insert a new column named Category where you label
entries as Data Book if the title contains the word Data; otherwise,
use Other.
```

In my case, Copilot generates two possible solutions. The first is the following formula to apply the conditional logic manually:

```
=IF(ISNUMBER(SEARCH("Data",[@Title])),"Data Book","Other")
```

To ask Copilot to explain the previous formula, use the following prompt:

```
Explain the following formula:
=IF(ISNUMBER(SEARCH("Data",[@Title])),"Data Book","Other")
```

In my case, Copilot generated the following explanation:

```
This formula categorizes book titles based on whether they contain the word
"Data." Here's how it works:
  • Formula: =IF(ISNUMBER(SEARCH("Data",[@Title])),"Data Book","Other")
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

- Purpose: Checks if the word "Data" exists in the "Title" column of the current row in the table and assigns a category accordingly.

Explanation:

1. SEARCH("Data",[@Title]):
   - The SEARCH function looks for the word "Data" in the "Title" column of the current row.
   - If "Data" is found, it returns the position of the first occurrence (a number). If not found, it returns an error.
2. ISNUMBER(SEARCH("Data",[@Title])):
   - The ISNUMBER function checks if the result of SEARCH is a number (indicating "Data" was found).
   - Returns TRUE if "Data" is found, otherwise FALSE.
3. IF(ISNUMBER(...),"Data Book","Other"):
   - The IF function evaluates the result of ISNUMBER:
     - If TRUE (word "Data" is found), it assigns "Data Book."
     - If FALSE (word "Data" is not found), it assigns "Other."

Result:

- Titles containing "Data" are labeled as "Data Book."
- Titles without "Data" are labeled as "Other."

The second solution provided by Copilot is a preview table in the chat box, shown in Figure 7-8.

Figure 7-8. The preview table provided by Copilot in the chat box

Click the "Insert column" button to insert a new column in the original table.

The prompt you just saw is an example of applying conditional logic to text. Table 7-3 outlines additional types of conditional text logic, along with common use cases and

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

example prompts you can use with Copilot.

Table 7-3. Examples of conditional text logic with use cases and sample Copilot prompts

Logic type	Possible use case	Prompt to Copilot
Tagging by category or keyword	Assign a label based on specific values in a text field	Consider the Title column. Insert a new column named Category where you label entries as Data Book if the title contains the word Data; otherwise, use Other.
Detecting presence or absence	Identify whether a string contains or lacks expected elements	Consider the Title column. Insert a new column named HasStorytelling where you check if the word Storytelling is present in the text. Return Yes or No accordingly.
Highlighting data	Mark entries that are empty, partial, incorrectly formatted, or match a given criterion	Consider the Title column. Highlight the entries in green if the word Storytelling is present in the text.

The table outlines three common types of conditional text logic: tagging by keyword, detecting presence or absence, and highlighting data based on criteria. These examples can serve as templates when you want to apply some conditional logic to text entries in your own datasets.

Summary

In this chapter, you explored a range of techniques for transforming text with Copilot. By now, you're likely starting to see the benefits, and may even be feeling a bit more confident and curious about what else Copilot can do:

Copilot can explain Excel formulas in plain language, helping users understand even complex logic.

You can use Copilot to generate descriptions of what each formula does, including its inputs and outputs. Use the story structure to better formulate

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

prompts for Copilot.

Sometimes a formula does not work well for specific data, and errors may occur. Ask Copilot to adapt the formula for the specific context where it is applied.

Use Copilot to formulate conditional text logic when you want to assign labels, tags, or outcomes based on the contents of a cell.

In the next chapter, you'll learn how to use Copilot to work with basic statistics.

Activity

Solve the cryptogram in Figure 7-9 using the numbers already decoded in the keyword. Use the known values to help decode the remaining letters. You can find the solution in Appendix D.

Figure 7-9. Cryptogram to solve

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Chapter 8. Basic Statistics with Copilot

I've always been passionate about cooking. Whether it's sweet pastries or savory dishes, I like experimenting in the kitchen. But as every cook knows, having ingredients alone isn't enough. You need the right proportions, the right timing, and the proper methods to bring them together into something worth serving.

Statistics works much the same way. Data, on its own, is just a collection of raw ingredients. Without the structure of statistical methods, the "recipes" that tell you how to measure, mix, and interpret, you can't transform those ingredients into something useful.

In this chapter, we'll explore how basic statistical thinking, paired with the capabilities of Copilot, can help you go from a jumble of raw numbers to clear, actionable insights. We will focus on descriptive statistics, but you can easily adapt the concepts described in this chapter to other types of statistics.

This chapter covers:

- An introduction to basic statistics
- Measures of central tendency
- Measures of dispersion
- Data distribution and shapes
- How to interpret statistical results

Let's start with introducing basic statistics.

Introducing Basic Statistics

Basic statistics are typically grouped into three core categories:

Measures of central tendency

> These summarize the center of a dataset using a single representative value.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Measures of dispersion

These quantify how much the data varies or spreads around that center.

Data distribution and shape

These describe the overall structure or profile of the dataset.

You'll explore each of these categories in detail throughout this chapter. For now, it's enough to understand their roles and how they complement one another in describing a dataset.

To calculate any of these metrics using Copilot in Excel, it helps to follow a simple prompt pattern based on two elements:

The context identifies the data column you're working with.

The main point specifies the statistical measure to compute.

As noted earlier, select the Copilot modality according to your goal: Agent Mode for making changes to the worksheet, Analysis Mode for calculations that shouldn't alter the data, and chat for general guidance. The key is not the feature name, which may change, but the intent behind how you use it.

For example, you can write the following prompt:

```
Consider column X Calculate measure Y using an Excel formula.
```

Typically, Copilot suggests inserting the result below column X as an output. In the next section, you'll see how this structure can be applied in practice, starting with measures of central tendency.

Measures of Central Tendency

Measures of central tendency describe the center of a dataset using a single value. The most commonly used metrics are mean, median, and mode. Each has its own strengths and is best suited for specific data scenarios. Table 8-1 describes each measure, indicates when to use it, and provides the Excel formula to calculate it.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 8-1. Measures of central tendency

Name	Description	When to use	Excel formula
Mean (average)	The sum of all values in a dataset divided by the number of values	The data is symmetrically distributed, with no outliers	=AVERAGE(A2:A100)
Median	The middle value in an ordered dataset	Skewed data or when outliers are present, as it is not affected by extreme values	=MEDIAN(A2:A100)
Mode	The value that occurs most frequently in a dataset	For identifying repeated values in a numeric dataset	=MODE.SNGL(A2:A100)

TIP

If your data is multimodal, MODE.SNGL will return only the first mode encountered in the dataset. Use MODE.MULT instead when you want all modes returned.

To understand how to use Copilot to calculate central tendency measures, consider the Datasaurus Dozen dataset created by Justin Matejka and George Fitzmaurice, inspired by the Datasaurus set from Alberto Cairo. A simplified version of the dataset is available in the GitHub repository for the book at 08/datasaurus.xlsx. Figure 8-1 shows an extract of the dataset.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 8-1. An extract of the Datasaurus Dozen dataset

The dataset contains three worksheets (Dino, Star, and Bullseye). Each worksheet contains a table with the following columns:

- x: The x coordinate
- y: The y coordinate

Open the file in Excel, select the Dino worksheet, and access the Copilot chat box by selecting it from the Home tab of the Ribbon. Select cell A1 and write the following prompt to calculate the average value:

```
Consider column X Calculate the mean using an Excel formula.
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

In my case, Copilot provided the following output:

```
Mean of x: 54.26327324
Looking at A1:B143 on 'Dino', here's 1 formula to review and insert in cell A144:
=AVERAGE(Dino[x])
```

In some cases, Copilot can also provide the Insert button, which inserts the output directly under the column and outside the table.

Proceed as follows to insert the formula in the current worksheet:

1. Select cell H2 and write the text Measure.
2. Select cell I2 and write the text x.
3. Select cell H3 and write the text Mean.
4. Select cell I3 and paste the formula =AVERAGE(Dino[x]) proposed by Copilot.

Figure 8-2 shows the result.

Figure 8-2. The table with the first measure for column x

Now, repeat for column y. Select cell A1 in the Dino worksheet and write the following prompt for Copilot:

```
Consider column Y. Calculate the mean using an Excel formula.
```

In my case, Copilot suggests the following formula:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
=AVERAGE(Dino[y])
```

Proceed as follows to add the measure to the table in Figure 8-3:

1. Select cell J2 and write the text y.
2. Select cell J3 and paste the formula =AVERAGE(Dino[y]) proposed by Copilot.

Figure 8-3 shows the resulting table.

Figure 8-3. The table with the first measure for columns x and y

Just as you calculated the average (mean), you can also compute the median and mode to gain a better understanding of your data. Use the following prompt with Copilot to calculate the median value:

```
Consider column x. Calculate the median using an Excel formula.
```

Select cell H4 and write the text Median. Select cell J4 and paste the formula =MEDIAN(Dino[x]) proposed by Copilot. Repeat for column y.

Use the following prompt with Copilot to calculate the mode:

```
Consider column x. Calculate the mode using an Excel formula.
```

Select cell H5 and write the text Mode. Select cell J5 and paste the formula =MEDIAN(Dino[x]) proposed by Copilot. Repeat for column y.

Now select all the cells from H2:H5 to J2:J5. Then, go to the Home tab on the Ribbon and click Format as Table. Choose your preferred table style from the options, and click

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

OK in the dialog box that appears. Click the Table Design tab in the Ribbon and write DinoStats in the text box on the upper-left part of the Ribbon. This will be the name for this table. Figure 8-4 displays the final table, which includes all the measures.

Figure 8-4. The table with all the measures for columns x and y

Now ask Copilot to draw a scatter plot with all the data. Select cell A1 and write the following prompt in the Copilot box:

```
Draw a scatter plot for Table Dino.
```

Copilot generates the chart in Figure 8-5. Click the ”Add to sheet” button to add the chart to the current sheet.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 8-5. The scatter plot of the Dino dataset

The chart is incredible! It shows a T. rex. My son Antonio would be excited to see it since he likes them very much.

Replicating Stats Across Worksheets

Now, we want to repeat the same steps we applied to the Dino worksheet for each of the other worksheets as well. Ask Copilot to list the steps to follow by writing the following prompt:

> I want to copy the DinoStats table to each worksheet, but have it reflect the specific data from that worksheet instead of the original. Suggest how I can perform this task.

In my case, Copilot provided the following answer:

> To copy the "DinoStats" table to each worksheet and have it reflect the specific data from that worksheet, follow these steps:
>
> Steps:
> 1. Copy the "DinoStats" Table:
>   - Select the "DinoStats" table (H2:J5 on the "Dino" sheet).
>   - Copy it (Ctrl+C or right-click > Copy).
> 2. Paste the Table on Other Worksheets:
>   - Go to the "Star" worksheet.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

- Paste the table in the desired location (e.g., H2:J5).
- Repeat for the "Bulls eye" worksheet.

3. Update Formulas to Reference the Correct Data Table:
   - On the "Star" worksheet:
     - Update the formulas in the "x" column (I3:I5) to reference the "Star[x]" column.
     - Update the formulas in the "y" column (J3:J5) to reference the "Star[y]" column.
     - Example formulas:
       =AVERAGE(Star[x])
       =MEDIAN(Star[x])
       =MODE(Star[x])
       =AVERAGE(Star[y])
       =MEDIAN(Star[y])
       =MODE(Star[y])
   - On the "Bullseye" worksheet:
     - Update the formulas in the "x" column (I3:I5) to reference the "Bullseye[x]" column.
     - Update the formulas in the "y" column (J3:J5) to reference the "Bullseye[y]" column.
     - Example formulas:
       =AVERAGE(Bullseye[x])
       =MEDIAN(Bullseye[x])
       =MODE(Bullseye[x])
       =AVERAGE(Bullseye[y])
       =MEDIAN(Bullseye[y])
       =MODE(Bullseye[y])
4. Verify the Results:
   - Ensure the formulas are correctly calculating the statistics based on the respective worksheet's data table.

By following these steps, the "DinoStats" table will reflect the specific data from each worksheet.

WARNING

Sometimes Copilot may suggest the old MODE function. Don't use it. MODE is a legacy function kept only for backward compatibility. The modern, supported equivalent is MODE.SNGL, which returns a single mode (the most frequent value). Always use MODE.SNGL in new formulas to avoid issues in current and future versions of Excel.

Follow the steps described by Copilot. The three datasets all have the same mean, but they look completely different. Note that the mode formulas for the Star and Bullseye tables return #N/A because they have no modes. To verify this visual diversity, select the Star worksheet and ask Copilot to draw the chart, using the following prompt:

Draw a scatter plot for Table Star.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Click the “Add to sheet” button. Figure 8-6 shows the resulting chart.

Figure 8-6. The scatter plot of the Star dataset

Do the same for the Bullseye dataset. Write the following prompt for Copilot:

```
Draw a scatter plot for Table Bullseye.
```

Click the “Add to sheet” button. Figure 8-7 shows the resulting chart.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 8-7. The scatter plot of the Bullseye dataset

TIP

The examples we've explored clearly show that measures of central tendency, like the mean, aren't enough to fully capture the shape or distribution of a dataset. That's why I strongly recommend always visualizing your data: a chart can reveal patterns, outliers, and structures that summary statistics alone might miss.

Using a Single Prompt for All the Measures

So far, you have calculated all the measures independently and manually built a table. You can combine all the commands and ask Copilot to generate everything in a single shot. Select cell A1 in the Dino table. Then, write the following prompt:

```
Consider columns x and y in the Dino worksheet. Calculate the mean, median, and
mode using an Excel formula, and list the instructions to build a table with
the following columns: Measure, x, and y.
```

In my case, Copilot answered with all the formulas and the following instructions to build the table:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Here are Excel formulas to calculate mean, median, and mode for columns x and y in the Dino table:

```
=AVERAGE(Dino[x])
=AVERAGE(Dino[y])
=MEDIAN(Dino[x])
=MEDIAN(Dino[y])
=MODE(Dino[x])
=MODE(Dino[y])
```

Instructions to build the table:

1. Create a new table with columns: Measure, x, y.
2. In the Measure column, enter: Mean, Median, Mode.
3. In the x column, enter the formulas above for x (mean, median, mode).
4. In the y column, enter the formulas above for y (mean, median, mode).

Copilot also proposed inserting the resulting table in a new worksheet, as shown in Figure 8-8.

Figure 8-8. The table suggested by Copilot for insertion as a new worksheet

Click the “Insert to new sheet” button. Figure 8-9 shows the resulting table.

Figure 8-9. The resulting table after clicking the “Insert to new sheet” button

A few weeks ago, I was planning a trip to the beach with my family. I checked the weather forecast and saw that over the coming days, temperatures would range from 24°C to 29°C, with an average of about 25.5°C (75.2°F to 84.2°F, with an average of

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

about 77.9°F). If I had only looked at that average, I might have decided to skip the beach entirely. At that temperature, the water would probably feel too cold for a swim.

But instead, I looked closer. I noticed that on some days the temperature would reach 29°C, which is warm enough for a comfortable swim. We chose those warmer days for our trip, and sure enough, the water felt just right.

That's exactly why measures of central tendency alone aren't enough. The average gives you the center of the data, but without knowing how much the values vary, you miss opportunities and make less informed decisions. Dispersion tells you the rest of the story.

Measures of Dispersion

While measures of central tendency describe the center of a dataset, measures of dispersion describe how spread out the data is around that center. Dispersion reveals whether values are tightly grouped or widely scattered, which is critical for interpreting the reliability and variability of your data. The most common measures of dispersion include range, variance, standard deviation, and interquartile range (IQR). Table 8-2 describes each measure, indicates when to use it, and provides the Excel formula to calculate it.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 8-2. Measures of dispersion

Name	Description	When to use	Excel formula
Range	The difference between the highest and lowest values in a dataset	When you need a quick, rough estimate of the data spread	=MAX(A2:A100) - MIN(A2:A100)
Variance	The average of the squared differences from the mean	To understand the variability of the dataset	=VAR.S(A2:A100)
Standard deviation	The square root of the variance. It shows how much values deviate from the mean.	To assess how spread out the values are around the mean	=STDEV.S(A2:A100)
IQR	The difference between the 75th and 25th percentiles (Q3 – Q1)	When dealing with skewed data or outliers	=QUARTILE.EXC(A2:A100, 3) - QUARTILE.EXC(A2:A100, 1)

The process of guiding Copilot to calculate dispersion measures closely resembles the steps used for central tendency measures. However, to better understand how to use Copilot in practice for calculating dispersion, we'll walk through a brief example. For more detailed strategies, refer to the previous section.

Consider 08/datasaurus.xlsx again. You can find details about the worksheets and tables in "Measures of Central Tendency". Write the following prompt for Copilot to calculate all the measures of dispersion together:

> Consider columns x and y in the worksheet, Dino. Calculate the range, the variance, the standard deviation, and the IQR using an Excel formula, and list the instructions to build a table with the following columns: Measure, x, and y.

In my case, Copilot provided the following answer:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Here's how to calculate Range, Variance, Standard Deviation, and IQR for columns x and y in the Dino worksheet using Excel formulas:

Formulas

Range

```
=MAX(Dino[x]) - MIN(Dino[x])
=MAX(Dino[y]) - MIN(Dino[y])
```

Variance

```
=VAR.S(Dino[x])
=VAR.S(Dino[y])
```

Standard Deviation

```
=STDEV.S(Dino[x])
=STDEV.S(Dino[y])
```

Interquartile Range (IQR)

```
=QUARTILE.INC(Dino[x], 3) - QUARTILE.INC(Dino[x], 1)
=QUARTILE.INC(Dino[y], 3) - QUARTILE.INC(Dino[y], 1)
```

Instructions to Build the Table

1. Create a table with columns: Measure, x, y.
2. In the Measure column, enter: Range, Variance, Standard Deviation, IQR.
3. In the x column, enter the formulas above for x.
4. In the y column, enter the formulas above for y.

WARNING

Copilot may suggest using QUARTILE.INC, but be careful. QUARTILE.INC includes the minimum and maximum values when computing quartiles, which matches older Excel behavior but can distort results, especially with small datasets. In contrast, QUARTILE.EXC excludes the extremes, following the statistical definition of quartiles. For most analytical or academic work, QUARTILE.EXC provides a more accurate representation of data distribution. Use QUARTILE.EXC unless you specifically need backward compatibility.

Copilot also proposed inserting the resulting table in a new worksheet, as shown in Figure 8-10.

Figure 8-10. The table suggested by Copilot for insertion as a new worksheet

Click the "Insert to new sheet" button. Figure 8-11 shows the resulting table.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 8-11. The resulting table after clicking the “Insert to new sheet” button

Visualizing Measures of Dispersion

You can use Copilot to draw charts related to measures of dispersion. Use the following basic prompt to draw a specific measure:

```
Create a histogram of the data in column x of table Dino to visualize the
frequency distribution.
```

Replace the word histogram with the type of chart you want to build and the sentence to visualize the frequency distribution with the content you want to show. In this example, Copilot generates the chart in Figure 8-12. Click the “Add to sheet” button to add the chart to the current sheet.

Figure 8-12. The histogram generated by Copilot for the Dino table

In some cases, Copilot can’t draw the chart directly. In this case, rephrase the prompt to ask for instructions, for example:

```
Suggest the steps to build a box plot for column x in table Dino
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

In my case, Copilot provided the following answer:

Here are the steps to build a box plot for column x in the table Dino:

1. Select the Data:
   - Click anywhere in the Dino table.
   - Select both columns "x" and "y" (Dino[x] and Dino[y]).
2. Insert a Box Plot:
   - Go to the Insert tab.
   - Click on Insert Statistic Chart (box-and-whisker icon).
   - Choose Box and Whisker.
3. Adjust the Chart Data (if needed):
   - If the chart includes both columns, right-click the chart and select Select Data.
   - Ensure only "x" and "y" are included as series.
4. Format the Chart:
   - Add chart and axis titles as needed.
   - Right-click the box plot to format outlier markers, whiskers, etc.
5. Interpret the Chart:
   - The box shows the interquartile range (Q1 to Q3).
   - The line inside the box is the median.
   - Whiskers show the range, and dots indicate outliers.

Figure 8-13 shows the resulting chart after following the steps described by Copilot. In Step 4, I also added the legend, although Copilot's instructions were more generic.

Figure 8-13. The box plot for the Dino table, drawn from the suggestions provided by Copilot

Exercise 1

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Consider 08/datasaurus.xlsx again. Use Copilot to calculate the measures of dispersion for the Star and Bullseye worksheets. Can you identify any similarities in the dispersion measures across the different tables?

You can find the solution to the exercise in 08/exercise1_solution.txt.

Data Distribution and Shape

Understanding a dataset's shape is essential to interpreting how data behaves beyond its central tendency and dispersion. Skewness and kurtosis are two advanced metrics that quantify shape. Table 8-3 describes each measure, indicates when to use it, and provides the Excel formula to calculate it.

Table 8-3. Measures of data distribution and shape

Name	Description	When to use	Excel formula
Skewness	Measures the asymmetry of the data distribution	When you want to assess whether data is symmetrically distributed or skewed	=SKEW(A2:A100)
Kurtosis	Measures the "tailedness" of the data distribution	When analyzing how heavily data tails differ from a normal distribution	=KURT(A2:A100)

Skewness measures the asymmetry of a distribution. A skewness of 0 indicates a perfectly symmetrical distribution. A positive skew means a longer right tail (outliers on the high end). A negative skew means a longer left tail (outliers on the low end). Figure 8-14 helps clarify this aspect.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 8-14. The different types of skews

Kurtosis measures the “tailedness” or peakedness of a distribution. A kurtosis of 0 indicates a normal (mesokurtic) distribution. High kurtosis (>0) (leptokurtic) means a sharp peak and heavy tails. Low kurtosis (<0) (platykurtic) means a flatter peak and thinner tails. Figure 8-15 helps clarify the different types of kurtosis.

Figure 8-15. The different types of kurtosis

The steps for using Copilot to calculate skew and kurtosis are similar to those for calculating central tendency and dispersion. To see how this works in practice, we’ll

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

walk through a quick example. For more detailed guidance, refer to the previous section.

Consider 08/datasaurus.xlsx again. You can find details about the worksheets and tables in "Measures of Central Tendency". Write the following prompt for Copilot to calculate all the measures of data distribution and shape together:

```
Consider columns x and y in the worksheet, Dino. Calculate the skewness and
kurtosis using an Excel formula, and list the instructions to build a table
with the following columns: Measure, x, and y.
```

In my case, Copilot provided an answer similar to the previous two examples. It provided the steps to build the table, a preview of the table, and the button to add it to the worksheet, as shown in Figure 8-16.

Figure 8-16. The table about skew and kurtosis suggested by Copilot for insertion as a new worksheet

In addition to the previously generated content, Copilot suggested some possible prompts to explain measures and interpret the results, as shown in Figure 8-17. This is precisely the topic of the following section.

Figure 8-17. The prompts suggested by Copilot to explain measures and interpret the results

So far, you have seen how to use Copilot to calculate measures of tendency, dispersion, distribution, and shape. In the next section, you'll see how to use Copilot to interpret the calculated measures.

Interpreting Statistical Results

Copilot can help you make sense of the results, suggest next steps, and even summarize insights right inside your spreadsheet.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

In this section, we'll walk through how to use Copilot to interpret statistical results using a simplified version of the Datasaurus dataset, found in the book's GitHub repository at 08/datasaurus_statistics.xlsx.

This workbook includes two sheets:

Dino: The Dino data table, with x and y columns

Statistics: All the key measures already calculated, a result of giving Copilot the following prompt:

```
Consider columns x and y in the worksheet, Dino. Calculate the mean, median,
mode, range, the variance, the standard deviation, the IQR, skew and kurtosis
using an Excel formula, and list the instructions to build a table with the
following columns: Measure, x, and y.
```

As with previous examples, we'll build our Copilot prompt around two key ingredients:

Context

Which worksheets, tables, and/or columns to include in the analysis

Main point

The specific type of analysis or interpretation we want Copilot to deliver

In the following pages, we'll walk through some example questions that show you how to use Copilot to interpret statistics. By following the same prompt structure, you'll be able to craft your own variations and have Copilot uncover additional insights tailored to your dataset.

Reliability of the Calculated Measures

You can ask Copilot to compare the mean with the other values in the dataset. This helps you to understand if there are outliers or if the dataset is skewed. In both cases, consider using the median value instead of the mean to better describe the dataset.

Consider the 08/datasaurus_statistics.xlsx dataset. Open it in Excel and access the Copilot chat box from the Home tab of the Ribbon. Select cell A1 of the table in the Dino worksheet and write the following prompt for Copilot:

```
Consider table Dino. Given the values of the mean and median for column x,
analyze which measure more accurately reflects the data distribution, and
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
explain what this reveals about the presence of skewness or outliers in
the dataset.
```

The prompt defines the following:

> The context, which specifies the table to consider and the column in the table: `Consider table Dino. Given the values of the mean and median for column x`
>
> The main point, which specifies the analysis to perform: `analyze which measure more accurately reflects the data distribution, and explain what this reveals about the presence of skewness or outliers in the dataset.`

In my case, Copilot provided the following answer:

```
The average of x is 54.26327323943664 and the median is 53.3333, which are very
close, indicating a roughly symmetric distribution with little skewness or
outliers.
```

According to the answer provided by Copilot, the mean and median are pretty similar, so both measures describe the dataset well. In this case, the answer is obvious and anchored to real, calculated measures, so you can trust it. There are some cases, however, where Copilot may hallucinate and provide wrong interpretations. For this reason, I suggest using the interpretations provided by Copilot as preliminary and quick ways to understand your data.

In addition, to make the answers provided by Copilot more robust and with fewer hallucinations, I suggest applying the following two strategies:

Provide the same prompt as input to Copilot multiple times

> If you always obtain the same data interpretation, the probability that the answer is correct is higher. For example, I provided the same prompt to Copilot for a second time and obtained the following answer, which is slightly more detailed than the previous one:
>
> ```
> For column "x" in table Dino, the mean is 54.26 and the median is 53.33
> (from the Statistics table). These values are very close, indicating that
> the data distribution for "x" is approximately symmetric and not heavily
> skewed. The skewness value (0.29) also supports this, as it is close to
> ```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

zero.

In such cases, both the mean and median accurately reflect the center of the data. The small difference suggests there are no significant outliers or strong skewness affecting the distribution. Thus, either measure is appropriate, but the mean is slightly more sensitive to any potential outliers. In this dataset, the similarity between mean and median shows outliers are minimal or not influential.

Reformulate the question using different words

For example, you can reformulate the prompt related to the reliability of the calculated measures as follows.

In the 'Statistics' table, for column x, the mean is listed under 'Mean' and the median under 'Median'. Explain the difference between these two values and what this difference suggests about the distribution.

In my case, Copilot answered:

In the 'Statistics' table for column x, the mean (54.26) is slightly higher than the median (53.33). This difference suggests the distribution of x is mildly right-skewed (positively skewed), meaning there are some higher values that pull the mean above the median. In a perfectly symmetrical distribution, the mean and median would be equal.

This answer is similar to the answer resulting from the initial prompt, meaning that the provided answer is quite reliable.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

WARNING

In AI, a hallucination occurs when the model produces output that sounds plausible but is factually incorrect or entirely made up. Hallucinations occur because large language models generate text by predicting likely word sequences, not by verifying truth. When using AI in critical tasks, always validate results against trusted sources instead of assuming they are correct.

Presence of a Dominant Value

You can ask Copilot to analyze the distribution of values in a dataset to determine whether there is a single dominant value or multiple values. Write the following prompt for Copilot:

```
Consider Table Dino. Analyze the distribution of values in column x and determine
whether the dataset is unimodal, bimodal, or multimodal. Identify the mode(s) and
explain what this reveals about possible subgroups or patterns in the data.
```

As in the previous examples, the prompt defines:

- The context, which specifies the table to use
- The main point, which details the type of analysis to perform

In my case, Copilot responded by creating the chart shown in Figure 8-12, preceded by the following text:

```
The distribution of values in column x is shown above; you can identify the
mode(s) and assess whether the dataset is unimodal, bimodal, or multimodal
from the frequency chart.
```

In this response, Copilot didn't interpret the measures directly. For this reason, I wrote another prompt for Copilot:

```
Explain the steps to identify if the dataset is unimodal, bimodal, or multimodal,
looking at the chart.
```

Note that in this case, I didn't add the context to the prompt because I formulated a generic question, unrelated to the specific data. Copilot provided the following answer:

```
To identify if the dataset is unimodal, bimodal, or multimodal using a chart
(like a histogram or frequency plot), follow these steps:
  1. Create a frequency chart: Plot the values from the "x" column of Table Dino
     on the x-axis and their frequencies on the y-axis.
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

2. Observe the peaks: Look for distinct peaks (local maxima) in the chart.
3. Count the modes:
   - Unimodal: One clear peak.
   - Bimodal: Two distinct peaks.
   - Multimodal: More than two peaks.
4. Interpret the modes: Each peak suggests a concentration of values, possibly indicating subgroups or patterns in the data.

In summary, the number of peaks in the frequency chart tells you if the distribution is unimodal, bimodal, or multimodal.

Looking at the data distribution in Figure 8-12, we can conclude that there is just one peak, so the dataset is unimodal in column x.

Exercise 2

Consider the 08/datasaurus_statistics.xlsx file. Use Copilot to do the following:

1. Compare the mean and the median values for column y.
2. Determine if there is a single dominant value or multiple values in column y.

You can find the solution to this exercise in the book's GitHub repository at 08/exercise2_solution.txt.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

LEARNING TO SWIM

I'd like to end this chapter with a memory that has stayed with me. My daughter Giulia had a deep longing to learn how to swim. She started with a bright inflatable ring that kept her safely afloat, letting her feel the water without fear. Over time, she traded the ring for a pair of arm floaties, gaining more freedom to move. Then, one radiant summer morning, she slipped them off entirely and swam like a fish, gliding freely through the open sea.

I stood there, watching her cut through the water with confidence, and I thought: learning to use AI is so much the same. At first, you need that ring, the security of foundational strategies, which I've been sharing with you throughout this book, and I'll continue doing so until the end. Those strategies keep you afloat while you find your balance. Then come the arm floaties: practical examples that let you try your strokes, steering, and adjusting, without the fear of sinking.

And then, one day, you're ready. You take the floaties off. You dive in. You begin to chart your own course, navigating the vast and shimmering sea of AI, sometimes drifting with the currents, sometimes setting your own heading, confident that you can go as far as your curiosity will carry you.

Summary

In this chapter, you have learned how to use Copilot to calculate and interpret basic statistics:

Basic statistics include measures of central tendency, dispersion, distribution, and shape. Measures of central tendency (mean, median, mode) describe the center of a dataset. Measures of dispersion (range, variance, standard deviation, IQR) quantify data variability. Measures of distribution and shape (skewness, kurtosis) help you understand the structure of the dataset.

To use Copilot to quickly calculate these measures, follow a prompt pattern that combines context (columns and tables to analyze) and main point (the statistical measure to compute).

You can use Copilot to create visualizations (scatter plots, histograms, box plots) to reveal shapes, patterns, and outliers that summary statistics alone may miss.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

You can use Copilot to interpret results by asking targeted questions to assess data symmetry, identify dominant values, and recognize whether a distribution is unimodal, bimodal, or multimodal.

To improve the reliability of Copilot's interpretations, rephrase or repeat questions. This strategy helps Copilot reduce the risk of incorrect or hallucinated answers.

In the next chapter, you'll learn how to perform data analysis using Copilot.

Activity

Solve the crossword puzzle shown in Figure 8-18. The solution is in Appendix D.

Figure 8-18. Crossword to solve

Across

2. The middle value in a dataset when the numbers are arranged in order

6. The Excel formula for a measure of the "tailedness" of a data distribution compared to the normal distribution

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

9. Abbreviation for operating system, the software that manages computer hardware and basic functions

11. Abbreviation for Estimated Number Output

12. Abbreviation for Data Table

13. Abbreviation for Number Type

14. Abbreviation for Work In Worksheet

15. Abbreviation for error, as in Excel error codes (for example, #DIV/0!)

18. A total or sum of values in a dataset

21. Abbreviation for number

22. The smallest unit of a spreadsheet where data is entered

23. A structured collection of related data

26 . Microsoft's spreadsheet program used for data analysis and visualization

Down

1. The formula name for a measure of the asymmetry of a data distribution

3. To pull or retrieve specific data from a dataset or source

4. Abbreviation for insight, as in actionable insights from data

5. Prefix meaning not, used in terms like nonparametric

7. A standard measure used for data values

8. A horizontal arrangement of cells in a spreadsheet

10. Short for statistics

12. Abbreviation for Data Extraction

16. A defined instruction or formula applied to data

17. The value occurring most often in a dataset

19. The average value of a dataset

20. Abbreviation for Ordinary Least Squares, a method for linear regression

24. Abbreviation for algorithm

25. Abbreviation for Trend Analysis

Chapter 9. Data Analysis with Copilot

This past summer, I was out in the countryside with my family, picking blackberries. We do this every year: we eat some straight off the brambles, and we bring the rest home. One year I even tried making jam, though I'll admit it wasn't much of a success.

As I was picking, something struck me. I wasn't gathering every berry in sight. I was selective, taking only the plumpest and ripest ones. And that's when the analogy hit me: data analysis is a lot like berry picking. You don't use every piece of data; you focus on the best data, the parts of the dataset that lead to meaningful insights.

Think of the brambles as a table with many columns. The berries are like the columns you care about. The skill lies in filtering, in choosing which data to focus on so that the important signals stand out from the noise. That's exactly why, in this chapter, we'll explore techniques for building PivotTables, which is a way of separating the ripe berries from the rest.

But there's more to it. When you're in the field, you notice that some berries aren't ready yet; you need to come back a day or two later. Data can be similar. Sometimes the answer isn't directly present in the raw numbers. Through What-If Analysis, you can simulate scenarios and see how outcomes might change if certain conditions were different.

Of course, no single chapter can cover every data analysis technique that Excel provides. My goal here is to give you a foundation, a way of thinking and working with Copilot, that you can build on and adapt to your own needs and datasets.

This chapter covers:

- Advanced PivotTables
- What-If Analysis
- Generalizing data analysis

Let's start with advanced PivotTables.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Advanced PivotTables

Advanced PivotTables in Excel are one of the most powerful tools for exploring and analyzing large datasets. Unlike simple summaries, PivotTables allow users to dynamically rearrange data fields and instantly change perspectives without altering the underlying dataset.

In Chapter 5, you learned how to build PivotTables in Excel using Copilot. As a quick recap, you can structure the prompt for Copilot using the narrative-based strategy: context, main point, and action. More precisely, structure the prompt as follows:

Context

> Specify the dataset or focus (for example, "Consider the MainTable table").

Main point

> Describe what analysis or summary you want (for example, "total revenue by category").

Action

> Add how you want the result displayed or adjusted (for instance, "sort descending").

The third part of the prompt (Action) is optional, and you can remove it if you don't want any specific format.

Throughout this chapter, for demonstration purposes, we'll use a dataset representing restaurant orders collected over one year. The dataset is available in the GitHub repository for the book at 09/restaurant_orders.xlsx.

The dataset includes the following fields:

Order ID

> A unique identifier for each order line (e.g., ORD10001)

Order Date

> The date the order was placed (YYYY-MM-DD)

Dish

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

The name of the ordered dish (e.g., Tiramisu, Margherita Pizza)

Category

The dish category, such as Starter, Main Course, Second Course, Dessert, Beverage, or Tasting Menu

Unit Price

The price of a single portion of the dish, in dollars. Includes variations due to inflation and promotions

Quantity

The number of portions ordered for that dish in the specific order line

Waiter

The name of the server who took the order (e.g., Luca, Anna, John)

Table

The table number (from 1 to 20) assigned to the order, relevant for dine-in service

Payment Method

How the bill was paid, either Cash or Card

Channel

The sales channel: Dine-in, Takeaway, or Delivery

Revenue

Calculated field: Unit Price × Quantity, representing the total value of the order line

Year

The year of the order (e.g., 2023 or 2024)

Month

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

The month of the order, as a number from 1 to 12; useful for time-based analysis

Figure 9-1 shows a screenshot of the dataset.

Figure 9-1. A screenshot of the restaurant_orders dataset

The next two subsections discuss dynamic summaries and slicers, which are important tools in data analysis.

Creating Dynamic Summaries

Dynamic summaries allow you to explore and reorganize large datasets without changing the underlying data. Through dynamic summaries, you can instantly switch perspectives, compare categories, analyze trends over time, and filter data interactively.

To understand how to work with Copilot to structure a dynamic summary, we'll use the restaurant_orders dataset available at 09/restaurant_orders.xlsx. We'll use this dataset to build a dynamic view of revenue by category, month, and sales channel. To get started, open the dataset in Excel and access the Copilot chat box by selecting it in the Home tab of the Ribbon.

As discussed earlier, choose the Copilot modality based on your objective: use Agent Mode to modify the worksheet, Analysis Mode to perform calculations without changing the data, and the standard chat for general questions or guidance.

First, we'll create a summary that shows which categories contribute the most to total revenue. Let's write the following prompt for Copilot:

```
Consider the Orders table. Create a PivotTable that shows total Revenue by
Category and sort the values in descending order.
```

Figure 9-2 shows the output produced by Copilot in my case.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 9-2. The preview of the PivotTable generated by Copilot

Click the “Add to a new sheet” button to insert the PivotTable as a new sheet.

TIP

In some cases, Copilot does not generate the PivotTable. If that happens, simply rerun the same prompt. In other cases, Copilot may answer that it can’t order the values in the table. If so, replace the previous prompt with the following one:

```
Consider the Orders table. Create a PivotTable that shows
total Revenue by Category. Describe the steps to sort
the values in descending order.
```

Now let’s extend the PivotTable with a monthly breakdown of revenue to reveal seasonal trends. Go back to the Orders worksheet and select any cell within the Orders table. Then write the following prompt for Copilot:

```
Consider the current PivotTable. Add Month as a column from the Orders table so
it shows monthly totals by Category.
```

Copilot generates a new PivotTable and suggests adding it as a new sheet, per the usual “Add to a new sheet” button. After clicking this button, you should see an output similar

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

to that shown in Figure 9-3.

Figure 9-3. The PivotTable showing the monthly revenue by category

Just for a quick check, click a cell inside the PivotTable (for example, B4) and select the PivotTable tab from the Ribbon. Click the Field List button to show the PivotTable fields (Figure 9-4). In some cases, if you click a cell in the PivotTable, it automatically displays the PivotTable Fields pane on the right.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 9-4. The pane with the PivotTable fields.

You can use this pane to configure the fields involved in the PivotTable. The selected fields are shown as headings in the PivotTable (Category, Revenue, and Month, as shown in the example). For example, if you want to show the monthly revenue by Waiter, select the Waiter checkbox and deselect Category. You should obtain a new PivotTable, as shown in Figure 9-5.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 9-5. The PivotTable showing the monthly revenue by waiter

Adding Slicers

Slicers enable you to visually filter a PivotTable by clicking buttons instead of using drop-down menus. They make dashboards more intuitive and reader-friendly.

NOTE

Slicers are currently available only in the desktop version of Excel. In Excel for the web, Copilot can help generate formulas or summary tables that simulate similar results. For each feature discussed in this section, I'll describe both the desktop functionality and a suggested workaround for the web version.

Consider the restaurant orders dataset available at 09/restaurant_orders.xlsx again. We will enhance the PivotTable created in the previous section (the one in Figure 9-5) with a slicer, allowing us to focus on sales by waiter or by channel. Go back to the Orders worksheet, click a cell in the PivotTable, and give Copilot the following prompt:

```
Consider the current PivotTable. Add slicers for Category and Channel to make the
PivotTable interactive.
```

In my case, Copilot provided the following steps to add slicers:

```
To add slicers for Category and Channel to your current PivotTable (PivotTable2
on Sheet2):
  1. Click anywhere inside PivotTable2.
  2. Go to the PivotTable Analyze tab on the Ribbon.
  3. Click Insert Slicer.
  4. In the Insert Slicers dialog, check Category and Channel.
  5. Click OK
```

Figure 9-6 shows the result after following the steps described by Copilot. You can play with the Category and Channel slides to enable specific categories and channels.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 9-6. The PivotTable with the Category and Channel slicers enabled

If you use the web version of Excel, write the following prompt for Copilot:

```
I'm using the web version of Excel. Propose a workaround for this: Consider the
current PivotTable. Add slicers for Category and Channel to make the
PivotTable interactive.
```

The first part of the prompt ("I'm using…a workaround for this") defines a preliminary context that should help Copilot understand the environment's general settings. The second part of the prompt is the same one used for the desktop version of Excel, which supports slicers.

TIP

Include your environment's general settings as preliminary context in the prompt whenever you need Copilot to adjust for features your version of Excel doesn't support.

In my case, Copilot provided the following output:

```
Excel for the Web does not currently support adding slicers to PivotTables. As a
workaround, you can use filter dropdowns in the PivotTable fields or create
interactive summary formulas. Here's how you can achieve similar interactivity:
Option 1: Use PivotTable Filters
  1. Add "Category" and "Channel" as filters in your PivotTable:
     • Click anywhere in the PivotTable.
     • Open the PivotTable field list.
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
    • Drag "Category" and "Channel" to the Filters area.
  2. Use the filter dropdowns above the PivotTable to select the values you want.
Option 2: Create Interactive Summary with Formulas
  1. Reserve two cells (e.g., N1 for Category, O1 for Channel) for user
     selection.
  2. Use Data Validation to create dropdowns for unique "Category" and "Channel"
     values.
  . Use this formula to summarize Revenue based on selections:
     =SUMIFS(Orders[Revenue], Orders[Category], N1, Orders[Channel], O1)
This will update the summary based on the selected Category and Channel.
```

Figure 9-7 shows the result after applying option 1.

Figure 9-7. The PivotTable with the Category and Channel filters enabled

The second option provided by Copilot is slightly more complex, and I don't recommend applying it. Instead, I suggest using option 1. However, I'll use option 2 to show how you can use Copilot as a teacher that can guide you step-by-step to reach a specific goal.

TIP

If you need more assistance or are unsure about something, you can ask Copilot to delve into a specific topic or provide steps to follow for implementation. Copilot is your personal teacher, although you still have to double-check its instructions!

The first point in option 2 suggests reserving two cells for user selection. It does not specify the worksheet to use, so we'll create a new worksheet. Click the + button to add a new sheet. In addition, option 2 suggests using cells N1 for Category and N2 for Channel. Let's keep this in mind.

The second step of option 2 assumes that we are already familiar with the concept of data validation. If you know how to proceed, simply apply what Copilot is suggesting. Otherwise, you can ask Copilot to provide more details about something it's proposing. In this case, the following prompt asks for clarification regarding step 2:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
How can I use Data Validation to create dropdowns with unique values from
"Orders[Category]" and "Orders[Channel]"?
```

In my case, Copilot provided the following output:

```
To create dropdowns with unique values from
"Orders[Category]" and "Orders[Channel]" using Data Validation:
  1. Choose a helper area (e.g., another sheet) to extract unique values.
  2. Enter these formulas to get unique lists:
     For unique Categories:
     =UNIQUE(Orders[Category])

     For unique Channels:
     =UNIQUE(Orders[Channel])
  3. Select the cell where you want the dropdown.
  4. Go to Data > Data Validation.
  5. Choose "List" and set the Source to the range with your unique values
     (e.g., if your unique categories are in X2:X10, enter =$X$2:$X$10).
  6. Click OK.
Now your dropdown will show only unique values from the selected column.
```

Figure 9-8 shows the result after following the steps described by Copilot. As a helper area, I have used cell A2 for Category and B2 for Channel. In addition, I have used N1 and O1 for user selection, as previously suggested by Copilot. Compared to the steps described by Copilot, I have added two labels in cells N2 and O2 to identify the Category and the Channel, respectively.

Figure 9-8. The result after applying the steps provided by Copilot to use data validation to create dropdowns

Now we can apply the third step of option 2. Paste the formula suggested by Copilot in cell P1. Figure 9-9 shows the result after selecting Dessert in the Category drop-down menu and Takeaway in the Channel menu. I have also added the label Sumin cell P2.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 9-9. The result after selecting options for Category and Channel

WARNING

Often, it's not worth using the exact cells suggested by Copilot, because the result is not very satisfactory, as shown in Figure 9-9, where I had to add labels below the cells to make it clearer what each individual cell corresponded to.

Of course, it's impossible for this chapter to cover every advanced technique for building PivotTables in Excel with Copilot. What I've tried to show you in this section isn't a catalog of features, but a way of thinking about how to work with Copilot.

TIP

My approach is simple: don't focus on memorizing every possible trick. Instead, focus on how you frame your prompts. As long as you stick to the principle of Context + Main Point, you'll be able to tackle almost any challenge your data presents. Over time, you'll discover that you can build increasingly sophisticated PivotTables, not because you know every feature in advance, but because you've trained yourself to reason through problems step-by-step.

If you remember, a few chapters back I talked about a dataset on maritime traffic and how I had to clean it before I could even start analyzing it. Back then (this was way back in 2017) Copilot didn't exist, so I had to perform every step of the analysis by hand. Looking back, I realize how much easier it would have been to simply ask Copilot:

```
Suggest PivotTables that make sense of this dataset.
```

What really matters is developing that mindset. I've had plenty of moments where I didn't know how to proceed, but simply explaining my problem to Copilot got me

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

unstuck. In my experience, Copilot will almost always propose something useful to get you moving again.

What-If Analysis

What-If Analysis in Excel allows you to simulate scenarios by changing key variables to observe how the results are affected. Excel provides two main What-If Analysis tools: Goal Seek and data tables. We won't focus on the Scenario Manager and Solver add-ins, but you can also apply the same logic explained here to use them.

We'll use the restaurant orders dataset to show how you can use Copilot to implement What-If Analysis. Similar to the other examples, we'll structure each prompt with two main elements:

> Context: Specifies which table or element to consider
>
> Main point: Defines what we want to do

Let's start with Goal Seek.

Goal Seek

Goal Seek determines the necessary input value to achieve a desired outcome. You specify a formula cell and set a target value, and Excel calculates the input that makes the formula return that value.

Consider the 09/restaurant_orders.xlsx dataset again. Open it and select the Orders worksheet. Let's suppose that you want to know how many portions of Spaghetti Carbonara must be sold to reach €1,000 in revenue. Write the following prompt for Copilot:

```
Consider the Orders table. Use Goal Seek to calculate how many portions of
Spaghetti Carbonara must be sold to reach €1,000 in revenue.
```

The prompt first defines the context ("Consider the Orders table") and then the main point (the second part of the prompt). In my case, Copilot didn't apply Goal Seek directly. Instead, it specified the steps to implement it, with the following output:

```
To solve this with Goal Seek, you need a cell that calculates total revenue for
"Spaghetti Carbonara" based on a variable quantity. Here's how you can set
it up:
Approach:
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

1. Calculate the average unit price for "Spaghetti Carbonara" from the Orders table.
2. Create a cell for the variable quantity (let's say in cell X1).
3. Create a formula to calculate revenue: Quantity × Average Unit Price.
4. Use Goal Seek to set the revenue formula to €1,000 by changing the quantity cell.

Step-by-step formulas:

1. Calculate average unit price for Spaghetti Carbonara:
   =AVERAGEIFS(Orders[Unit Price], Orders[Dish], "Spaghetti Carbonara")
2. Assume cell X1 is your variable quantity (enter any starting value, e.g., 1).
3. Calculate revenue in cell Y1:
   =X1 * AVERAGEIFS(Orders[Unit Price], Orders[Dish], "Spaghetti Carbonara")
4. Use Goal Seek:
   - Set cell Y1 to value 1000 by changing cell X1.

This will give you the number of portions of Spaghetti Carbonara needed to reach €1,000 in revenue.

Figure 9-10 shows the result of Copilot's suggested steps 1 to 3. Although the result is correct, that's not immediately clear. In fact, if you reopen the worksheet after a few months, I doubt that you'll remember what cells X1 and Y1 represent. For this reason, I suggest modifying the cells used for calculations by Copilot based on your specific needs. Also, remember to adjust the formulas based on your changes.

Figure 9-10. The result of steps 1 to 3 suggested by Copilot

You might not know how to implement Copilot's suggested step 4 if you are unfamiliar with Goal Seek. For this reason, I suggest asking Copilot to explain it by writing the following prompt:

```
Explain how to use Goal Seek
```

In my case, Copilot provided the following answer:

Here's how to use Goal Seek in Excel:

1. Enter a formula in a cell (e.g., F12) that depends on another cell (the input).
2. Go to the Data tab, then click What-If Analysis > Goal Seek.
3. In the Goal Seek dialog:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

- Set cell: Select the cell with the formula (e.g., F12).
- To value: Enter the result you want.
- By changing cell: Select the input cell that the formula depends on.

4. Click OK. Excel will adjust the input cell until the formula cell reaches your target value.

Goal Seek works best when the formula cell is directly or indirectly linked to the input cell.

In this example, the cell F12 proposed by Copilot corresponds to Y1, the value is 1000, and the input cell that the formula depends on is X1. Figure 9-11 shows how to fill in the Goal Seek dialog.

Figure 9-11. The parameters used to fill in the Goal Seek dialog

After clicking OK, you should see the output shown in Figure 9-12.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

A2.

Figure 9-12. The result of Goal Seek

Both cells, X1 and Y1, have been adjusted to the desired value: Y1 has been set to 1000 (the desired value) and X1 to 84.2204, which is the number of portions of Spaghetti Carbonara that must be sold to reach €1,000 in revenue.

The result of Goal Seek shown in Figure 9-12 is very ambiguous because it's not clear what each cell represents. For this reason, I suggest using a new sheet for the Goal Seek calculations and formatting the cells used for calculations as a table, as shown in Figure 9-13. This structure also enables you to add other dishes.

Figure 9-13. The reorganized table

Data Tables

Data tables allow you to evaluate how results change when varying one or two inputs:

A one-variable table shows the effect of changing a single input.

A two-variable table shows the combined effect of two inputs.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Consider the 09/restaurant_orders.xlsx dataset again. Open it and select the Orders worksheet. Let's suppose you want to test how changes in the Tasting Menu price (from €30 to €50) affect revenue. Tasting Menu is a particular value in the Category column.

Open the Copilot chat box and write the following prompt:

```
Consider the Orders Table. Create a one-variable data table showing how Tasting
Menu revenue changes as price varies from €30 to €50.
```

In my case, Copilot provided the following prompt:

```
To create a one-variable data table showing how "Tasting Menu" revenue changes as
price varies from €30 to €50, follow these steps:
High-level approach:
You'll set up a column with prices from 30 to 50, and next to it, use a formula
to calculate total revenue for "Tasting Menu" if that price is used for all
such orders. Then, use Excel's Data Table feature to automate recalculation.
Step-by-step:
  1. List prices:
     In a column (e.g., Y2:Y22), enter values from 30 to 50 (increment by 1 or
     as desired).
  2. Set up the revenue formula:
     Suppose Z1 contains the label "Revenue". In Z2, enter this formula to
     calculate total revenue for "Tasting Menu" if the price in Y2 is used
     for all such orders:
     =SUMIFS(Orders[Quantity], Orders[Dish], "Tasting Menu") * Y2
  3. Create the data table:
     • Select the range Y1:Z22 (including headers).
     • Go to Data > What-If Analysis > Data Table.
     • For "Column input cell", select any single cell (e.g., AA1) not used
       elsewhere (it won't affect the calculation since the formula
       references Y2 directly).
     • Click OK.
Result:
You'll get a table showing how total "Tasting Menu" revenue changes as the price
varies from €30 to €50.
```

The output provided by Copilot is not correct for step 3. In fact, you can't just select any single cell. You must select exactly the cell referencing the price variable. This example demonstrates that the output provided by Copilot and AI in general is not 100% guaranteed. One way to check if the generated output is correct is to proceed as follows:

1. Try to follow the steps proposed by Copilot. If you don't get any errors, verify that the final output is reasonable.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

2. If you encounter errors, ask Copilot if the procedure is correct.

To better understand the second step (asking Copilot if the procedure is correct), consider the previous output generated by Copilot. Step 3 (Create the data table) is not correct for bullet point three (For "Column input cell", select any single cell). If you try to apply this procedure by selecting, for example, cell AA2, you should obtain the error shown in Figure 9-14.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 9-14. The erroneous output obtained by applying Copilot's suggested procedure

As you can see in the figure, the value in cells Z2:Z22 is always the same and does not depend on the price. To solve this problem, ask Copilot the following prompt:

```
Is the following set of instructions correct? Set Up Data Table: Select the range
Y1:Z22 (Y1: "Price", Z1: "Revenue", Y2:Y22: prices, Z2: formula). Go to Data >
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

> What-If Analysis > Data Table. For "Column input cell", select any cell (e.g., Orders[Unit Price][first Tasting Menu row])—the formula uses Y2, so this is just a placeholder.

In my case, Copilot provided the following answer:

> Your instructions are mostly correct, but there is one important correction:
> For the "Column input cell," you must select the actual cell that the formula in Z2 references for the price variable (e.g., the cell in your model that represents the Tasting Menu price). It should not be just any cell or a placeholder; it must be the cell whose value you want to vary in the data table.
> Summary:
> - Select Y1:Z22 (Y1: "Price", Z1: "Revenue", Y2:Y22: price values, Z2: formula referencing the price cell).
> - Go to Data > What-If Analysis > Data Table.
> - For "Column input cell," select the cell that the formula in Z2 uses for price (not just any cell).
>
> Let me know if you want a step-by-step guide or help with the formula setup.

Copilot clarifies that we must use the cell that the formula in Z2 uses for price (not just any cell), which corresponds to Y2 in our case. Figure 9-15 shows the correct output.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 9-15. The correct output for a one-variable data table

Exercise

So far, you have learned how to build a one-variable table, which shows the effect of changing a single input. Now you can use the same approach to build a two-variable

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

table to show the combined effect of two inputs.

Consider the 09/restaurant_orders.xlsx dataset again. Write a prompt asking Copilot to create a two-variable data table showing revenue changes when both the price (€30–50) and quantity sold (10–30 units) vary.

Figure 9-16 shows the output. You can find a possible prompt for Copilot in 09/exercise1_solution.txt.

Figure 9-16. The two-variable data table

Generalizing Data Analysis

So far, you've explored some of the tools that Copilot offers for data analysis in Excel, including advanced PivotTables and What-If Analysis. These are powerful features, but they come with an assumption: you already know what kind of analysis you want to perform.

In practice, this assumption doesn't always hold. You might be working with a dataset you don't fully understand, or you may simply not know what analytical approaches are most relevant. In such cases, you can turn the problem on its head and let Copilot itself suggest what kinds of analyses make sense.

The strategy is straightforward:

Ask Copilot to propose analyses

> Provide a simple prompt that asks Copilot to suggest types of analysis suited to your data.

Repeat the request

> Run the same prompt two or three times to check for consistency in Copilot's suggestions.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Curate the results

From the suggested list, select only those analyses that seem appropriate for your goals.

Delegate execution

Ask Copilot to carry out the chosen analysis on your behalf.

Review carefully

Validate the outputs before proceeding to the next analysis type.

Let's walk through an example. Suppose you're working with the dataset 09/restaurant_orders.xlsx. Open Copilot and enter the following prompt:

```
Consider the Orders table. Suggest which types of analysis make sense for this
table.
```

Run this prompt again two more times. In my case, the results are summarized in Table 9-1.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 9-1. The results of running the same prompt across three runs

Run 1	Run 2	Run 3
Sales Analysis: Identify the most popular dishes and categories by quantity sold and revenue generated.	Sales Analysis: Identify top-selling dishes and categories by quantity and revenue.	Sales Analysis Total revenue by month, year, dish, category, or channel. Average order value over time. Revenue trends (seasonality, growth).
Revenue Trends: Examine how revenue changes over time (daily, monthly, yearly) to spot seasonality or growth patterns.	Revenue Trends: Analyze revenue over time (daily, monthly, yearly) to spot seasonality or growth.	Product Performance Best-selling dishes or categories. Revenue and quantity sold per dish/category. Menu item popularity by channel (Dine-in, Delivery, Takeaway).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Run 1	Run 2	Run 3
Waiter and Table Performance: Assess which waiters and tables generate the most revenue or handle the most orders.	Waiter and Table Performance: Assess which waiters and tables generate the most revenue or handle the most orders.	Staff Performance Revenue and number of orders handled by each waiter. Average order value per waiter.
Channel and Payment Method Analysis: Compare dine-in, takeaway, and delivery performance, and see which payment methods are most used.	Channel and Payment Method Analysis: Compare performance across dine-in, takeaway, and delivery, and analyze payment preferences.	Customer Behavior Popular payment methods. Table usage patterns (which tables are most/least used).
Customer Behavior: Explore order combinations, average order value, and peak ordering times.	Customer Behavior: Explore order combinations, average order value, and peak ordering times.	Operational Insights Order volume by day, week, or month. Channel performance comparison.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Run 1	Run 2	Run 3
		Profitability Analysis
		Revenue per order, per dish, or per category.
		High/low margin items (if cost data is available).
		Time-Based Analysis
		Trends by year and month.
		Peak times/days for orders.
		Custom Segmentation
		Filtering by specific waiters, dishes, or payment methods for targeted analysis.

The first two runs produced the same types of analysis, while the third run was more detailed.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Suppose you want to proceed with the sales analysis. Write the following prompt for Copilot:

```
Consider the Orders table. Identify the most popular dishes and categories by
quantity sold and revenue generated.
```

The prompt defines the context (Consider the Orders table), while the main point is exactly the text generated by Copilot in run 1 for sales analysis.

In my case, Copilot produced the results shown in Figure 9-17.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 9-17. The result of asking Copilot to perform sales analysis

As you can see from the figure, Copilot didn’t propose any formula or steps to calculate the results. Instead, it provided the results directly.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

DON'T TRUST COPILOT BLINDLY

When Copilot delivers a result without showing its work, treat it with caution. A single number or chart that appears out of nowhere may be correct, but you have no way of knowing unless you understand how it was produced.

Instead of accepting the output at face value, ask Copilot to explain the procedure it used. Request the intermediate steps, formulas, or logic that led to the final result. This way, you can validate the reasoning, confirm that the method is appropriate for your dataset, and spot potential mistakes before they propagate into your analysis.

Think of Copilot not just as a calculator that hands you answers, but as a partner whose reasoning you should always be able to inspect.

At this point, you can repeat the process for other analysis types that catch your attention. Each iteration builds a richer picture of your dataset, without requiring you to know all the right questions in advance.

Summary

In this chapter, you learned how to use Copilot within Excel to analyze data through three key tasks: building advanced PivotTables, performing What-If Analysis, and generalizing the data-analysis process. In the next chapter, you'll see how to visualize data using data reports and storytelling techniques. Here are some takeaways:

Advanced PivotTables enable you to build dynamic summaries and slicers that highlight the most relevant parts of your dataset.

What-If Analysis provides strategies to simulate scenarios with Goal Seek and data tables to understand how changes in inputs affect outcomes.

When you're not sure where to start, ask Copilot to suggest analyses, and then curate and validate the results.

Frame prompts with the context and main point, and always inspect Copilot's reasoning rather than trusting results blindly.

Activity

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Solve the word puzzle in Figure 9-18 by circling all the words from the list below. In the end, you will discover something about Copilot by reading the letters that are not circled, which will now form words. You can find the solution in Appendix D.

Figure 9-18. Word search to solve

Words to search for:

ACTION

CONTEXT

COPILOT

DATA

EXCEL

FILTER

GOAL SEEK

PIVOT TABLE

PROMPT

SLICER

WHAT IF

COPILOT

Word puzzle solution. Copilot: _____

Chapter 10. Data Visualization and Reporting with Copilot

Data visualization and reporting mark the final stage in the data analysis workflow. You start by cleaning your data, then analyzing it, and finally, you reach the visualization phase. Too often, though, by the time you get here, your energy or your schedule has run out. The result can be rushed, cluttered charts that don't do justice to all the hard work that came before.

You've invested hours scrubbing and exploring your data, yet the visuals at the end can look confusing or unfinished. I get it. Personally, visualization and reporting are my favorite parts of the entire data pipeline. This is where clarity comes into play. This is where you make everything you've done understandable to others and you invite them in. And that clarity can spark something powerful: wonder.

I remember the first time my kids rode a city bus. For me, it was routine because I've done it thousands of times. But for them, it was pure discovery. They stared out the window, wide-eyed, as if the whole world had gathered just outside that glass. That sense of wonder is exactly what you want to inspire in your audience.

You already know your data inside out because it's familiar, maybe even ordinary, to you. But for your audience, it's brand new. They're like my kids on that bus, seeing it all for the first time. Your job is to capture their attention, to make them curious, and to light that spark that makes them want to know more.

In this chapter, I'll show you how to use Copilot to build clear, engaging visualizations and reports quickly and effectively. Specifically, we'll cover the following:

- Data visualization
- Data reporting

But before we dive into the fun part, let's start by introducing the dataset we'll use throughout the chapter.

The Dataset

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

In this chapter, we'll use the dataset related to the yearly number of objects launched into outer space, released by Our World in Data under the CC BY 4.0 License[1]. You can also download the workbook containing the dataset from the GitHub repository of the book, under 10/space_objects.xlsx. The workbook contains a single worksheet with a table with the following four columns:

`Entity`: The name of a place

`Code`: ISO country code

`Year`: The reference year

`N`: Number of objects launched into outer space for that year

Figure 10-1 shows a snapshot of the table.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 10-1. A snapshot of the table in the space_objects.xlsx workbook

Data Visualization

Data visualization is turning raw data into a visual representation that makes patterns, trends, and insights easier to perceive. While tables and raw numbers may contain the truth, it is through charts and visualizations that we often discover the story hidden inside. Good data visualization is about making data clear, meaningful, and actionable.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

When using Copilot inside Excel, the way you frame your prompt determines the quality of the chart it generates. As I have already said throughout the book, a proper structure is to think of the prompt as a story with three parts:

Context

Define the data source. Which table, which columns, which filters?

Main point

Specify the visualization type and how the data should be mapped (for example, x-axis, y-axis, color categories).

Action

Add properties that refine the result: titles, axis labels, formatting choices, or removing extraneous elements.

To understand how you can apply this structure practically, consider the space_objects.xlsx dataset. As usual, I'll leave it to you to decide which Copilot modality best fits your goal: Agent Mode for modifying the worksheet, Analysis Mode for performing calculations without changing the data, and standard chat for general questions or guidance. What matters most is not the label, but your understanding of the intent behind each option. After opening the file in Excel and activating Copilot, write the following prompt:

```
Consider the Objects table. Insert a line chart showing N by year, using only
Entity = World. Add y axis label: "Number of Objects launched" and a chart
title, "Objects Launched in the World by Year".
```

Let's break this prompt down:

Context

`Consider the Objects table` tells Copilot where to look. It establishes the scope of the request.

Main point

`Insert a line chart showing N by year, using only Entity = World` specifies the visualization type (line chart), the mapping (N on the y-axis, year on the x-axis), and a filter condition (Entity = World).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Action

> `Add y axis label: “Number of Objects launched” and a chart title, “Objects Launched in the World by Year”` provides the finishing touches.

In my case, Copilot suggested adding the generated chart to a new sheet. Figure 10-2 shows the result generated by Copilot after clicking the button. Note that Copilot has also generated a PivotTable containing only the relevant data for the visualization.

Figure 10-2. The PivotTable and line chart generated by Copilot to show the number of objects launched in the world by year

When you build a chart, you should follow two foundational principles that guide effective data visualization:

- Choose the right chart type.
- Simplify and remove clutter.

In the rest of this section, we’ll see how to use Copilot to apply each of these principles in turn. Let’s start with choosing the correct chart.

Choose the Right Chart Type

The same dataset can be visualized in multiple ways, but selecting an effective chart type makes the difference between clarity and confusion. Some common guidelines can help you decide which chart to use as a starting point, based on the data you have.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 10-1 shows how to select an effective chart based on some popular objectives and data types. The table also describes the kind of insight you can extract using that specific chart, along with an example of a prompt to ask Copilot for implementing this chart.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 10-1. Popular charts based on the objective and data type

Objective	Data type	Chart	Insights	Prompt example
Show change over time	Time series	Line chart	Peaks, outliers, seasonality, trends, increases/decreases	Consider the cts table. In a line chart ing N by Year ing only Enti World.
Compare categories	Discrete categories	Column or bar chart	Ranking, magnitude differences, comparisons across groups	Consider the cts table. Se only Year = 2 Select only E y = United St s, Entity = C a, and Entity ussia. Insert olumn chart s ng N for Unit tates, China, Russia in 202
Show proportions	Parts of a whole	Pie or doughnut chart	Relative share, dominance, contribution to total	Consider the cts table. Se only Year = 2 Select only E y = United St s, Entity = C a, and Entity ussia. Insert ie chart show N by Entity f nited States, na, and Russi 2024.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Objective	Data type	Chart	Insights	Prompt example
Explore relationships	Two variables	X Y Scatter plot	Correlation, clusters, outliers, spread	`Consider the cts table. Se only Entity = ted States, E y = China. In an X Y scatte ot showing N Entity = Unit tates on the is and N for ty = China on y-axis.`

When working with Copilot, you should clearly indicate which chart type you want in the Main Point section of your prompt. The more explicit you are, the more likely you'll get exactly what you need. For example, you can write:

```
Insert a line chart.

Insert a bar chart.

Insert a pie chart.
```

However, you should always accompany the chart type with what you want to see on each axis or dimension. For instance, instead of just saying "Insert a line chart," say "Insert a line chart showing N for Entity = United States." This ensures the visualization reflects the story you want to tell.

WARNING

In some cases, Copilot doesn't produce the chart, but it provides an answer like the one shown in Figure 10-3. Usually, this means that you haven't properly instructed Copilot on how to filter the data. I suggest referring to the "Prompt example" column in Table 10-1 to learn how to properly filter data before creating a chart.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 10-3. The answer provided by Copilot in some cases

You may be unsure of the best visualization to use for your data. In this case, you can let Copilot choose for you by writing a generic instruction like “Insert a chart” without any specific detail about the chart type. Copilot will then attempt to suggest or generate the visualization type it finds most appropriate based on your dataset. For example, you might provide a prompt like this:

```
Consider the Objects table. Select only Year = 2024. Select only Entity = United
States, Entity = China, and Entity = Russia. Insert a chart showing N for United
States, China, and Russia in 2024.
```

In my case, Copilot generated the bar chart and the PivotTable shown in Figure 10-4.

Figure 10-4. The bar chart and PivotTable generated by Copilot

Sometimes you may want creative input without giving Copilot complete control. In that case, you can ask Copilot to suggest possible visual representations of your data. For example, you can write the following prompt:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

> Consider the Objects table. Select only Year = 2024. Select only Entity = United States, Entity = China, and Entity = Russia. Suggest 5 possible visual representations for comparing Entity types and explain the pros and cons of using each.

In my case, Copilot provided the following output:

> Here are five possible visual representations for comparing the number of objects launched by the United States, China, and Russia in 2024:
>
> 1. Bar Chart
>   - Pros: Easy to compare values side by side; clear for small numbers of categories.
>   - Cons: Can become cluttered with too many categories or small differences.
> 2. Pie Chart
>   - Pros: Good for showing proportions of a whole; visually intuitive for a few categories.
>   - Cons: Hard to compare similar-sized slices; not ideal for more than a few categories.
> 3. Horizontal Bar Chart
>   - Pros: Useful if entity names are long; easy to read and compare values.
>   - Cons: Takes up more vertical space; less common than vertical bar charts.
> 4. Stacked Bar Chart
>   - Pros: Shows both total and individual contributions; good for comparing parts to a whole.
>   - Cons: Harder to compare individual segments unless one is much larger.
> 5. Dot Plot
>   - Pros: Simple and clear for a small number of categories; easy to see differences.
>   - Cons: Not as visually impactful as bars; less familiar to some audiences.
>
> Each of these visualizations has strengths and weaknesses depending on what you want to emphasize—absolute numbers, proportions, or ease of comparison.

You now have more criteria to select the best chart for your data. Choose the chart type you like the most and ask Copilot to implement it, based on the guidelines given at the beginning of this section.

Simplify and Remove Clutter

One of the most common mistakes in data visualization is adding too much. Extra gridlines, redundant labels, decorative effects, and unnecessary series can distract from the story your data is trying to tell. Clutter doesn't just make charts ugly. It makes them harder to read and interpret.

Here are the most frequent sources of noise in Excel charts:

Too many gridlines

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Dense horizontal or vertical lines that compete with the data

Chart junk

3D effects, background images, or unnecessary shading

Over-labeling

Data labels on every point, repeated legends, or axis titles that restate the obvious

Too many categories

Pie charts with dozens of slices or bar charts with unreadable x-axis labels

Redundant series

Including entities that don't add to the analysis, such as "World" totals when you only care about individual countries

The principle is simple: remove anything that doesn't support the main message of your chart.

As of February 2026, Copilot can't directly edit or declutter charts you've already created. To reduce clutter, you have two options:

1. Generate a new chart with decluttering instructions.
2. Ask Copilot for the steps to declutter a chart manually.

In the first strategy, when you ask Copilot to create a chart, append decluttering instructions at the end of the end part of your prompt (for example, "Remove gridlines" or "Remove axis title"). This doesn't always work, not because Copilot can't follow the request, but because many decluttering tasks involve more complex adjustments.

Consider the dataset 10/space_objects.xlsx again. Select cell A1 in the main worksheet, and start Copilot. Use a structured prompt with the usual three parts (context, main point, action). In the end, add the decluttering step:

```
Consider the Objects table. Group by column Entity and select only Entity =
United States, Entity = China, and Entity = Russia. Insert a line chart
with a line for each selected Entity. The line chart shows N over Year. Add
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
y axis label: "Number of Objects launched". Add chart title: "Objects
Launched in the World by Year". Remove gridlines.
```

Notice the final instruction: "Remove gridlines." You can replace it with any other decluttering request ("Hide the legend," "Remove axis labels," and so on).

If you want to filter or reduce categories (for example, limiting the number of series in the chart), include that in the context section of the prompt, not in the decluttering action.

WARNING

In some cases, Copilot generates an image representing the chart rather than the chart itself. If this happens, try to rerun the prompt multiple times until Copilot generates the chart. If the problem persists, use the second option, described next.

Please note that the current version of Copilot doesn't support every type of decluttering. You'll need to experiment. If the request fails, try the second strategy.

In the second strategy, you ask Copilot for the steps to declutter manually. Here, you treat Copilot as a tutor. Instead of generating a new chart, you ask it for the exact steps to remove a specific element. For example, you can write the following prompt:

```
Describe the steps to remove gridlines in a chart.
```

Copilot will return a step-by-step procedure that you can apply directly to your chart. In my case, Copilot generated the following output:

```
How to Remove Gridlines from a Chart in Excel
  1. Click on the Chart. Select the chart from which you want to remove the
     gridlines.
  2. Open Chart Elements.
     a. Click the green "+" button (Chart Elements) that appears next to the
        upper-right corner of the chart.
     b. Alternatively, you can right-click on the gridlines directly in the chart.
  3. Uncheck "Gridlines".
     a. In the Chart Elements menu, uncheck the "Gridlines" option.
     b. If you right-clicked the gridlines, select "Delete" from the context menu.
  4. Gridlines will disappear. The chart will now display without gridlines,
     making it cleaner and less cluttered.
```

If you don't understand a specific step, you can ask Copilot to detail it. For example, you can write the following prompt:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

> Where do I find the chart's elements?

In my case, Copilot provided the following answer:

> In Excel, you can find the Chart Elements button as a small green "+" icon that appears next to the upper-right corner of your selected chart.
> Here's how to find and use it:
> 1. Click on your chart to select it.
> 2. Look for the green "+" icon (Chart Elements) that appears just outside the top-right corner of the chart border.
> 3. Click this icon to open a menu where you can add or remove elements like gridlines, axis titles, data labels, and more.
>
> Tip:
> If you don't see the green "+" icon, make sure the chart is selected. You can also access chart elements by right-clicking directly on parts of the chart (like gridlines or legend) to remove or format them

TIP

When you use Copilot as a tutor, you don't need the full context–main point–end structure, since the steps for decluttering in Excel are always the same regardless of the dataset.

Exercise 1

Consider the 10/sales.xlsx dataset available in the GitHub repository for the book. Use Copilot to perform the following tasks:

1. Insert a line chart showing NetUnits by Month.
2. Remove gridlines and set the title to Net Units Trend.

The solution to exercise 1 is available in 10/exercise1_solution.txt.

Data Reporting

Once you've created individual charts, the next step is to organize them into a report. A report can be interactive (such as a dashboard) or static (such as a summary). Reports guide the reader through insights from the data.

In Excel, you can organize a report as follows:

Summary

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

An introductory worksheet introducing the topic and the table of contents

Raw data

A worksheet with the raw data

Metrics and charts

One or more worksheets with the calculated metrics and their visual representation

Dashboard

A final worksheet with key performance indicators (KPIs) and summary data.

Let's see how you can use Copilot to help you generate each part of the report. We'll use the 10/space_objects_report.xlsx workbook, available in the GitHub repository for the book.

This workbook contains the following worksheets:

Raw Data

Objects by World

Timeline China Russia USA

Column Chart China Russia USA

Pie Chart China Russia USA

Bar Chart China Russia USA

In the remainder of this section, we'll see how to transform the workbook into a full report. We'll start with the Raw Data worksheet, then move on to defining the summary worksheet, and finally we'll build the dashboard. I won't cover the intermediate calculation sheets, since those were already discussed in the previous "Data Visualization" section.

Raw Data

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

This worksheet contains the raw data that was also included in the space_objects.xlsx workbook you used to build charts in the "Data Visualization" section.

To make this data easier to understand for your report's audience, it's a good idea to add the following two elements to the worksheet:

A legend describing the worksheet's columns

Some basic statistics, such as the total number of rows, the number of empty rows per column (if any), the years covered, and so on

If the raw data table is too large, consider adding this information in a separate worksheet. In our case, the table contains only four columns, so we'll include both the legend and the basic statistics in the same worksheet.

Let's start with the legend. Select the Raw Data worksheet and click cell A1. Then open Copilot and type the following prompt:

```
Consider the Objects table. Provide a description of the columns.
```

In my case, Copilot provided the following answer:

```
Here are the descriptions of the columns in the Objects table:
• Entity: The name of the country or organization associated with the object.
• Code: The three-letter country code (if available) for the entity, following
  international standards.
• Year: The year associated with the object, likely representing a launch or
  event year.
• N The count or number of objects/events for the given entity and year.
```

Copy the provided output, select cell F3, and paste the content. Adjust layout, font, and color, if required. Figure 10-5 shows my output, which includes a title. I used Arial font, size 15 for the title and 14 for the legend, dark blue font color, and bold for the title.

Figure 10-5. The legend in the Raw Data worksheet

Now, let's add some basic statistics to the Raw Data worksheet. Starting from cell F9, write the following text, one for each row, up to F14:

Basic Statistics

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Total Objects

Unique Entities

Non-Empty Code

Starting Year

Ending Year

Next, let's use Copilot to calculate the values for each basic statistic, starting with total objects. Write the following prompt for Copilot:

```
Consider the Objects table. Provide the formula to count the total number of
rows.
```

In my case, Copilot provided the following answer:

```
=ROWS(Objects[Entity])
```

Copy the preceding formula and paste it into cell G10.

Calculate the number of unique entities by writing the following prompt for Copilot:

```
Consider the Objects table. Provide the formula to count the number of unique
entities.
```

In my case, Copilot provided the following output:

```
=ROWS(UNIQUE(Objects[Entity]))
```

Copy the preceding formula and paste it in cell G11.

Repeat the operation for non-empty code, starting year, and ending year by writing, respectively, the following prompts:

> Consider the Objects table. Provide the formula to count the number of non-empty codes.
>
> Consider the Objects table. Provide the formula to calculate the min and max years.

In my case, Copilot provided the following formulas, respectively:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
=SUM(--(Objects[Code]<>""))
=MIN(Objects[Year])
=MAX(Objects[Year])
```

Copy the preceding formulas and paste them in cells G12:G14. Adjust the final layout, if necessary. Figure 10-6 shows a possible result.

Figure 10-6. The calculated basic statistics in the Raw Data worksheet

Once the Raw Data worksheet is ready, the next step is to add a Summary worksheet, which describes an overview of the content. The next section shows how to implement it.

Defining the Summary

The Summary worksheet introduces the report, providing readers with a quick overview of the workbook's purpose and structure. It usually contains the following:

- Title
- Overview of the topic
- Table of contents linking to the other worksheets

Open the space_objects_report.xlsx and create a new worksheet using the + symbol available to the right of all the worksheets. Drag the worksheet you just created to become the first worksheet in the series. Right-click the worksheet and select Rename from the menu. Rename the worksheet as Summary. Figure 10-7 shows the result.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 10-7. The Summary worksheet is the first worksheet

The next step is to populate this summary with the necessary information. Let's start with the title. We'll use Copilot to generate some possible titles for our report. Open Copilot and write the following prompt:

```
Propose five titles for a report showing statistics about the launch of objects
in outer space over the years.
```

In my case, Copilot proposed the following titles:

```
Here are five report title suggestions:
1. "A Historical Overview of Outer Space Object Launches: Trends and
   Statistics"
2. "Global Launch Activity: Entities, Codes, and Yearly Space Object
   Statistics"
3. "Tracking the Expansion: Annual Statistics of Space Object Launches by
   Entity"
4. "From 1957 to 2024: Key Metrics in the Launch of Outer Space Objects"
5. "Space Launches Across Decades: Entity Participation and Object Counts"
```

If you are not satisfied with these titles, you could ask Copilot to:

> Propose another five titles. Suggested prompt: "Propose another five titles."
>
> Elaborate on one of the previous titles. Suggested prompt: "Propose five alternative titles to 'From 1957 to 2024: Key Metrics in the Launch of Outer Space Objects.'"
>
> Propose other titles using a specific style. Suggested prompt: "Propose another five titles without using the colon."

TIP

Personally, I find the last strategy (the one that avoids using colons) the most effective. In general, AI systems (not just Copilot) tend to produce titles that include a colon, which often feel less natural or authentic. That's why I always recommend adding "without using colons" to your prompt when you're generating titles.

In my case, the title that stood out the most is "Key Metrics in Outer Space Object Launches from 1957 to 2024." Therefore, I'll use this title for our report example.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Select the Summary worksheet, then click on cell C2. Type the selected title into that cell. You can adjust the font size, color, and other formatting options to make it clear that this is the report's main title. For instance, I used Arial, 26 point, bold, and dark blue font, as shown in Figure 10-8.

Figure 10-8. The title of the report

The next step involves adding a short overview of the report. You can use Copilot to help you write it. First, ask Copilot to describe the columns of the table in the Raw Data worksheet. Next, ask Copilot to generate the short description, based on the columns.

For the short description, use the following prompt:

```
Based on the columns in the Objects table, provide a short description of the
dataset.
```

In my case, Copilot provided the following output:

```
This dataset records the annual launch activity of objects in outer space,
detailing the entity (country or organization) responsible, their
three-letter code (if available), the year of launch, and the count of
objects launched each year.
```

You can improve this basic text by asking Copilot to transform it into a report summary:

```
Modify your provided description so that it is the summary of a report where you
show statistics about the topic.
```

In my case, Copilot provided the following output:

```
This report presents key statistics on outer space object launches from 1957 to
2024. During this period, a total of 1,226 objects were recorded, involving 114
unique entities. Of these, 1,081 entries include a non-empty country or
organization code. The data spans launches from the earliest in 1957 to the
most recent in 2024.
```

Read the generated text carefully and modify it if necessary. Then copy the final text and paste it in cell C4 of the Summary worksheet. Adjust the font and the layout, if necessary. In my case, I pasted one sentence per row and increased the font size to 14, as shown in Figure 10-9.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 10-9. The title and the short description of the report

TIP

When you use Copilot to generate text, refine your prompts step-by-step. Start by asking it to write a general introduction to the topic. From there, progressively refine the request until you reach the level of detail and tone you want for the final description.

The last part of the Summary worksheet should contain the link to the other worksheets. We'll build a table containing all the worksheets. First, write the following prompt for Copilot:

```
List the worksheets in this workbook.
```

In my case, Copilot provided the following answer:

```
The worksheets in this workbook are:
  1. Summary
  2. Raw Data
  3. Objects by World
  4. Timeline China Russia USA
  5. Column Chart China Russia USA
  6. Pie Chart China Russia USA
  7. Bar Chart China Russia USA
```

Copy this list and paste it starting from cell C11. Write the following text in C10: Table of Contents. Optionally, adjust font size, color, and layout. Next, write the following prompt for Copilot:

```
Describe manual steps to create worksheet links
```

In my case, Copilot provided the following output:

```
To manually create worksheet links in Excel:
  1. Select the cell where you want the link.
  2. Right-click and choose "Link" (or "Hyperlink").
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

3. In the dialog, select "Place in This Document."
4. Choose the worksheet you want to link to from the list.
5. Optionally, enter a cell reference (e.g., A1) and a display name.
6. Click "OK."

Apply these steps for each worksheet, starting with cell C11, where you pasted the text summary. Figure 10-10 shows the final result.

Figure 10-10. The Summary worksheet

WARNING

You might consider asking Copilot to generate direct links to the worksheets—and that's exactly what I initially thought of doing, too. However, for security reasons, Copilot doesn't allow automatic generation of external links. The only option is to ask it for the steps to add those links manually. For example, I tried using the prompt: "Write the formulas to extract the link to each worksheet," but the response was: "A formula containing the HYPERLINK function won't be generated for this prompt to protect your privacy. Use another prompt and I'll try to help."

Dashboard

A dashboard is a single view that pulls together the most important metrics, visuals, and controls so you can monitor a topic at a glance. A good dashboard is usually interactive: readers can filter, slice, and drill into the data without touching the raw tables.

In this section, you'll see how to quickly build a simple dashboard in Excel using macros and Copilot. Recall from earlier in the book: an Excel macro is just recorded or written VBA code that automates a sequence of actions.

We'll assume you still have the workbook space_objects_report.xlsx with all the worksheets described earlier. We'll automatically create a new worksheet named

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Dashboard, copy all charts from the other worksheets into it, and add a slicer on the Entity field to filter the selected countries across all charts dynamically.

Open Copilot and write this prompt:

```
List the steps to copy all the charts in the various worksheets to a new
worksheet called Dashboard.
```

In my case, Copilot provided the following answer:

```
Here are the steps to copy all charts from the various worksheets to a new
worksheet called Dashboard:
  1. Create a new worksheet and name it Dashboard.
  2. Go to the first worksheet (e.g., "Objects by World") containing a chart.
  3. Click on the chart to select it.
  4. Press Ctrl+C to copy the chart.
  5. Go to the "Dashboard" worksheet.
  6. Click where you want to place the chart and press Ctrl+V to paste.
  7. Repeat steps 2–6 for each worksheet with a chart (Timeline China Russia USA,
     Column Chart China Russia USA, Pie Chart China Russia USA, Bar Chart China
     Russia USA).
  8. Arrange the charts as desired on the Dashboard worksheet.
```

Copy Copilot's step-by-step output and paste it into this follow-up prompt:

Write a macro that performs the following steps: [paste steps indicated by Copilot]

In my case, Copilot generated the following output:

```
CAUTION: Please carefully review the generated VBA code before running it!
Sub CopyAllChartsToDashboard()
    Dim wsNames As Variant
    Dim ws As Worksheet
    Dim dashWS As Worksheet
    Dim ch As ChartObject
    Dim i As Integer
    Dim leftPos As Double
    Dim topPos As Double

    ' List of worksheet names containing charts
    wsNames = Array("Objects by World", "Timeline China Russia USA", _
        "Column Chart China Russia USA", "Pie Chart China Russia USA", _
        "Bar Chart China Russia USA")

    ' Delete existing Dashboard if it exists
    Application.DisplayAlerts = False
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
    On Error Resume Next
    Worksheets("Dashboard").Delete
    On Error GoTo 0
    Application.DisplayAlerts = True

    ' Add new Dashboard worksheet
    Set dashWs = Worksheets.Add(After:=Worksheets(Worksheets.Count))
    dashWs.Name = "Dashboard"

    leftPos = 20
    topPos = 20

    ' Loop through each worksheet and copy its first chart
    For i = LBound(wsNames) To UBound(wsNames)
        Set ws = Worksheets(wsNames(i))
        If ws.ChartObjects.Count > 0 Then
            Set ch = ws.ChartObjects(1)
            ch.Copy
            dashWs.Paste
            ' Arrange charts in a grid: 2 per row
            With dashWs.ChartObjects(dashWs.ChartObjects.Count)
                .Left = leftPos
                .Top = topPos
                .Width = 400
                .Height = 250
            End With
            If (i + 1) Mod 2 = 0 Then
                leftPos = 20
                topPos = topPos + 270
            Else
                leftPos = leftPos + 420
            End If
        End If
    Next i
End Sub

You will need to use a macro-enabled workbook (.xlsm) to run this code.
```

Review Copilot's macro. If it looks reasonable, write the following prompt for Copilot:

```
Write another macro that adds a slicer in the Dashboard worksheet for the Entity
field synchronized with all the charts.
```

In my case, Copilot provided the following output:

```
CAUTION: Please carefully review the generated VBA code before running it!
Sub AddEntitySlicerToDashboard()
    Dim wsDashboard As Worksheet
    Dim ptList As Variant
    Dim pt As PivotTable
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
    Dim sc As SlicerCache
    Dim s As Slicer
    Dim i As Integer

    ' Set Dashboard worksheet
    Set wsDashboard = Worksheets("Dashboard")

    ' List of PivotTables to connect slicer to
    ptList = Array( _
        Worksheets("Objects by World").PivotTables("PivotTable4"), _
        Worksheets("Timeline China " & _
        "Russia USA").PivotTables("PivotTable11"), _
        Worksheets("Column Chart China " & _
        "Russia USA").PivotTables("PivotTable14"), _
        Worksheets("Pie Chart China Russia USA").PivotTables("PivotTable15"), _
        Worksheets("Bar Chart China Russia USA").PivotTables("PivotTable16") _
    )

    ' Delete existing slicer for Entity if it exists
    For Each sc In ThisWorkbook.SlicerCaches
        If sc.Name Like "Slicer_Entity*" Then
            sc.Slicers(1).Delete
        End If
    Next sc

    ' Add slicer for Entity to Dashboard, connected to the first PivotTable
    Set sc = ThisWorkbook.SlicerCaches.Add(ptList(0), "Entity")
    Set s = sc.Slicers.Add(wsDashboard, , "Slicer_Entity", "Entity", _
    20, 20, 200, 300)

    ' Connect slicer to all PivotTables
    For i = 1 To UBound(ptList)
        sc.PivotTables.AddPivotTable ptList(i)
    Next i
End Sub

You will need a macro-enabled workbook to run this code.
This macro adds a slicer for the "Entity" field to the "Dashboard"
worksheet and synchronizes it with all relevant PivotTables (and
thus all charts based on them).
```

Inspect the code. If it's correct for your workbook, create the two macros and run them. Here's a quick summary of how to make a macro in Excel (note that you have to run steps 1 to 3 only for the first macro):

1. On Windows: Go to File → Options → Customize Ribbon. In the right-hand panel, check the box next to Developer and click OK. On macOS: Click on the Excel menu in the top menu bar and select Preferences. Under the Authoring

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

section, click on Ribbon & Toolbar. In the Main Tabs list, check the Developer box and click Save.

2. On the Developer tab, click Visual Basic. Insert a new module: in the VBA editor, select Insert → Module. A blank code window appears.
3. Paste the VBA code generated by Copilot into the code window.
4. Back in Excel, save the file as a .xlsm file and then go to Developer → Macros, select your macro, and click Run.

You should see a new Dashboard sheet populated with copies of your charts and the slicer synchronized with all your charts (similar to Figure 10-11).

Figure 10-11. The dashboard generated through the macro

At this point, you can't use Copilot anymore because enabling macros automatically disables AutoSave. However, you can switch AutoSave back on (if it hasn't been switched on again automatically) and continue using Copilot. To do so, you'll need to remove the macro you just created before saving the workbook again. Here's how to do it:

1. Open the Visual Basic Editor.
    a. Go to the Developer tab and click Visual Basic, or press Alt + F11.
2. Locate the module that contains the macro.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

   a. In the Project Explorer (on the left), find your workbook (for example, VBAProject (space_objects_report.xlsm)).

   b. Expand it, then open the Modules folder.

   c. You'll see one or more modules (e.g., Module1).

3. Delete the module.

   a. Right-click the module that contains your macro and choose Remove Module1 (or whatever its name is).

   b. When prompted to export it, click No (unless you want to save a backup of the code).

4. Close the VBA editor.

   a. Click the X or press Alt + Q to return to Excel.

5. Save, close, and reopen the workbook.

   a. Save the file (you can keep it as .xlsm or save as .xlsx if you want to remove all macro functionality).

   b. Close Excel completely, then reopen the workbook.

After reopening, AutoSave will be re-enabled, and you'll be able to use Copilot again.

As a final step, add the Dashboard sheet to the Table of Contents in the Summary worksheet.

Exercise 2

Consider the 10/sales.xlsx dataset available in the GitHub repository for the book. Use Copilot to perform the following tasks:

Create a Summary worksheet.

Create a Dashboard worksheet.

The solution to the exercise is available in the GitHub repository for the book in 10/exercise2_solution.txt.

With this chapter, the second part of the book comes to an end. In Part II, you explored how to use Copilot in Excel, starting with data cleaning, proceeding through data

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

analysis, and ultimately reaching data visualization and reporting. In Appendix B, you'll find a collection of exercises related to this second part, written in a format I call applied fiction, a narrative, scenario-based way to practice what you've learned. I encourage you to take a look to see how it works in practice. Starting with the next chapter, we'll enter the third part of the book, where you'll discover other AI tools beyond ChatGPT (covered in Part I) and Copilot (covered in Part II).

Summary

This chapter covered the following:

- Copilot in Excel allows you to create visualizations, reports, and dashboards using natural language instead of complex commands or formulas. You describe what you need, and Copilot builds it for you.
- Effective use of Copilot follows a three-part prompt structure: context, main point, action. The context defines the data source and filters, the main point specifies the visualization type and data mapping, and the action adds formatting or refinement, such as titles and labels.
- Choosing the right chart is essential for clarity. Line charts show trends over time, bar or column charts compare categories, pie or doughnut charts display proportions, and scatter plots reveal relationships between variables. When you're unsure of which to use, Copilot can suggest chart types and explain their pros and cons.
- Simplification is a key principle. Removing unnecessary gridlines, labels, legends, or decorative effects makes charts easier to read. Clean visuals highlight the message behind the data. Copilot can provide step-by-step guidance to edit charts manually.
- Reports in Excel should be organized into structured worksheets, including a summary, raw data, charts, and a dashboard. Copilot helps generate each part by producing text, formulas, or layout suggestions.
- The Raw Data worksheet benefits from a legend that describes the columns and includes basic statistics, such as the total number of rows, unique entities, and the year range, all calculated using Copilot-generated formulas.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

The Summary worksheet introduces the report, providing a clear title, a concise description of the dataset, and a table of contents that links to the other worksheets. Titles can be generated by Copilot and refined iteratively to achieve the desired tone and style.

Dashboards consolidate key charts and indicators into a single, interactive view. Copilot can generate VBA macros that automate the process of copying charts from different worksheets and adding slicers for dynamic filtering. After running the macros, they should be removed to re-enable AutoSave and Copilot functionality.

Activity

Complete the following text using the keywords indicated below. The solution is in Appendix D.

COPILOT

DASHBOARD

EXCEL

LINE CHART

MACRO

REPORT

SUMMARY

VISUALIZATION

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

In this chapter, you learned how to use _____ in _____ to create a complete reporting workflow, from data cleaning to _____, building a _____, designing a _____, generating a _____, and even automating tasks with a _____, all to produce a clear and actionable _____.

1 Edouard Mathieu, Pablo Rosado, and Max Roser. (2022). "Space Exploration and Satellites". Published online at OurWorldinData.org.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Part III. Other AI Tools for Excel

Part III focuses on extending Excel beyond its native capabilities by integrating external AI tools and Python-based workflows. You'll learn how to connect Excel with specialized add-ins and programming tools bringing more automation, intelligence, and flexibility into your spreadsheets. The chapters show how Excel evolves from a traditional calculation tool into a hub for AI-assisted data processing, advanced analysis, and semi-automated workflows combining formulas, code, and external services.

Chapter 11 presents alternative AI tools to Copilot that can be used directly in Excel, many of which offer free plans that are especially suitable for working with small datasets. After introducing the role of add-ins and how they extend Excel with specialized AI capabilities, the chapter explores several tools in practice. You'll see how Numerous.ai brings AI into formulas for extraction and classification, how GPT for Work supports structured data extraction and text generation, how Power GPT assists with formulas and workflow guidance, and how Sally Suite enables AI-driven analysis and even provides Python code.

Chapter 12 shows how Python and AI can expand Excel into a more powerful and flexible analysis tool. We start with embedded Python in Excel using the `=PY()` function, learning how to run small analyses directly in cells and understanding its limitations. The chapter then introduces how to connect Excel, Python, and the OpenAI API, where Excel serves as the interface and Python handles processing and AI calls.

Appendix C includes activities inspired by the chapters in Part III, presented in the form of applied fiction. These narrative-based exercises allow you to practice technical concepts from the chapters in realistic, story-driven scenarios, reinforcing your understanding while keeping the learning experience engaging and creative.

Chapter 11. Landscape of Excel AI Add-Ins

When you first encounter AI in Excel, it's easy to think that one tool, specifically Copilot or ChatGPT, should be enough. After all, Copilot can summarize data, explain formulas, generate charts, and even help you write functions using plain language. So why would you ever need additional add-ins? If there's already an AI assistant built in, isn't everything else redundant?

As you begin to work with a real spreadsheet, which might include messy text fields, inconsistent product names, multilingual comments, nested formulas, and dashboards that need to be refreshed every month, you'll notice that not all AI capabilities are created equal. Copilot is excellent at helping you explore and understand. It's a conversational partner. You ask, it answers. You request, it responds. But Copilot isn't always optimized for the kind of controlled, reproducible operations you need in professional workbooks. In addition, Copilot comes with a cost.

That's where third-party AI add-ins come in. They don't replace Copilot; they specialize. Most of them provide a free plan: if you have a small dataset, you can manipulate it completely for free, using the free credit provided by the add-ins. Each add-in solves a different category of problem. Some add-ins bring AI directly into the grid, giving you formula-like functions that you can fill down just like `SUM()` or `VLOOKUP()`. Others automate end-to-end workflows: cleaning data, creating sheets, building charts, and preparing reports without requiring you to repeat the same steps over and over. Still others focus on helping you understand and improve your formulas, translating complex logic into readable English.

In this chapter, I'll walk you through the current landscape of AI add-ins for Excel, focusing on the following:

- Numerous.ai
- GPT for Excel Word
- Power GPT
- Sally Suite

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

But first I want to introduce the concept of add-ins in Excel.

Introducing Add-Ins in Excel

Add-ins are components that extend Excel with extra features. They can introduce new formulas, automate workflows, connect Excel to external services, or provide tools that are not available in the standard installation. Many add-ins are developed by third parties and are available through the Office Add-ins Marketplace.

Some add-ins connect Excel with cloud services, customer databases, or machine learning tools. The main advantage of using add-ins is that they allow you to keep working inside Excel while gaining access to advanced capabilities that would otherwise require switching to another application.

To install an add-in on Windows, click File > Account > Get Add-ins. If this option is disabled, select Home > Add-ins or Developer > Add-ins > Office Store. Enter the add-in name in the search field, select Add, accept the terms, and sign in if prompted. Alternatively, you can install an add-in directly from the Home tab of the Ribbon.

To install an add-in on macOS, click Insert > Add-ins > Get Add-ins. If this option is disabled, go to the Ribbon and select Home > Add-ins > More Add-ins > Office Add-ins or Developer > Add-ins > Office Add-ins. Search for the add-in name, select Add, accept the terms, and sign in if prompted.

AI-Powered Add-Ins

AI-powered add-ins bring language models and automated reasoning directly into Excel. They can interpret plain-text instructions, generate formulas, extract fields from long text values, or rewrite content stored in cells. These add-ins connect Excel to external AI services and execute requests from within the workbook. When you install an add-in, I recommend you also check its terms and privacy conditions.

Table 11-1 shows how AI-powered add-ins extend Excel with natural language processing and automated data transformations. They allow users to describe operations in plain language, apply those operations across worksheets, interact through an integrated chat panel, and use AI-driven functions directly in cells.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 11-1. Features provided by AI-powered add-ins

Feature	Description	Example
Natural language instructions	Allow users to describe the task in plain language instead of writing formulas or scripts.	Clean addresses, standardize labels, generate summaries, draft text content
Automated data processing	Applies the requested operation across multiple cells or the entire worksheet	Apply formatting rules, extract fields, transform text columns
Integrated chat panel	Includes a chat interface directly in Excel that can explain formulas, correct errors, or walk through changes step-by-step	Ask “Explain this formula,” request corrections, refine transformation logic
AI functions in cells	Provide custom Excel functions that work like standard formulas and can be filled down through cells	=NUMAI() =NUMINFER()

WARNING

AI-powered add-ins may require usage-based billing. Each operation can consume credits or tokens, depending on the provider. When working with large datasets, it is useful to start with small samples to monitor costs and accuracy before applying the process to the full dataset.

Comparing AI Add-ins and Copilot

AI add-ins and Microsoft 365 Copilot both bring language model capabilities into Excel, but they serve different workflows. AI add-ins usually focus on specific spreadsheet operations, while Copilot provides broader assistance across the entire Microsoft 365 environment. Understanding their differences helps determine which tool fits a particular task or organizational setup.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 11-2 compares AI add-ins with Copilot across key dimensions such as integration, cost, usage, and scope of use.

Table 11-2. Comparison between AI add-ins and Copilot

Feature	AI add-ins	Copilot
Integration	Work mainly inside Excel	Works across Microsoft 365 apps
Scope	Focused on specific spreadsheet tasks	Focused on specific spreadsheet tasks and broad assistance across documents and workflows
Usage	Function in cells or the chat panel inside the add-in	Unified natural language interface in Excel and other apps
Cost	Often usage-based (credits/tokens)	Requires a license per user
Best fit	When tasks are specialized and Excel-centered	When tasks are specialized and Excel-centered and when work spans multiple Microsoft 365 tools and workflows

In the remainder of this chapter, we will look at several add-ins, each offering different capabilities. Before we do that, let's check out the dataset we will use throughout the chapter.

Hospital Patient Dataset

The 11/patient.xlsx dataset is available in the GitHub repository for the book. This dataset represents a synthetic collection of hospital patient records. Each row corresponds to a single patient admission event and contains demographic attributes, clinical details, and administrative outcomes. Figure 11-1 shows a snapshot of the dataset.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 11-1. The hospital patient dataset

The table contains the following fields:

ID

A unique numerical identifier assigned sequentially to each record. It functions only as a reference key and is not tied to any real-world patient identification system.

Name

The patient's full name in FirstName LastName format.

Age

The age of the patient at the time of admission.

Sex

Indicates the patient's sex.

Diagnosis

The primary medical condition associated with the hospital stay. These descriptions are intentionally broad (for example, Influenza, Cancer, Migraine).

Department

The hospital unit or specialty responsible for the patient's treatment. Examples include Cardiology, Pediatrics, and Orthopedics.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Admission Date

The calendar date when the patient was admitted.

Discharge Date

The date when the patient left the care of the department.

Outcome

The status of the patient at the conclusion of the hospital stay. Possible values include Discharged, Under Treatment, Transferred, and Deceased, representing common end-of-care pathways.

Address

The patient's full residential address, entered as free text.

At this point, we are ready to work through practical examples using this dataset. We will begin with Numerous.ai.

Numerous.ai

Numerous.ai is an example of an add-in that brings AI directly into Excel formulas. Numerous.ai lets you use natural language inside formulas. Instead of writing complex text extraction, classification, or rewriting logic, you can simply describe what you want and refer to a cell or range. The add-in handles the rest and updates the results automatically in the data.

After installing the add-in, you must accept the terms of usage and sign in with your Microsoft or Google account. The service offers a seven-day free trial. After that, the base plan starts with a monthly token allowance. The number of tokens used depends on the text's length and the complexity of the request.

Numerous.ai provides the three main functions shown in Table 11-3.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 11-3. Functions provided by Numerous.ai

Function	Input parameters	Output	Purpose
NUMAI()	prompt, data_range	Transformed or extracted result based on the prompt	General AI text or data processing
NUMINFER()	examples_range, labels_range, new_value	Inferred label for the new value	Classification and pattern inference
NUMWRITE()	prompt, text_range	Rewritten, expanded, summarized, or translated text	Text generation and rewriting

The official documentation is very detailed, so I recommend referring to it for the full list of parameters and options. Here, I introduce how to use these functions in practical data scenarios.

Consider the 11/patients.xlsx dataset. One column contains complete addresses. Assume you want to extract only the city. You can use NUMAI() to do this:

1. Click cell K2.
2. Enter the following formula:

```
=NUMAI("Extract the city from", J2)
```

The first parameter is the prompt. The second parameter is the cell or range you want to process (in our case, the Address column). You can add more parameters if you want to provide additional instructions or reference other data.

After you press Enter, Numerous.ai begins processing. When it finishes, the extracted city names will appear in column K, as shown in Figure 11-2.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 11-2. Extracted city values in column K

Rename Column1 as City. Consider now the NUMINFER() function, which is designed for classification tasks. This function learns from examples you provide, then applies that pattern to new values. The function arguments are as follows:

```
NUMINFER(examples_range, labels_range, new_value)
```

examples_range is a range that contains sample inputs.

labels_range is a range with the classification labels for each sample input.

new_value is the cell you want to classify.

In other words, you show the model some examples of how you classify data, and it infers how you would classify new data.

Let's apply this to the patients.xlsx dataset. Add a new column named Priority to the dataset, as shown in Figure 11-3, and manually enter priority values for a few sample patients.

Figure 11-3. The new Priority column

To classify the rest automatically, in cell L7, I enter this:

```
=NUMINFER(E2:E6, L2:L6, [@Diagnosis])
```

Here is what is happening:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

E2:E6 is the example set of diagnoses.

L2:L6 contains the manually assigned priorities for those diagnoses.

[@Diagnosis] refers to the diagnosis in the current row. This is the value you want Numerous.ai to classify.

Once you press Enter, you see one result appear. To apply the function to the entire column, select the fill handle in the bottom right of cell L7 and drag it down to the end of the table.

Numerous.ai may display a “Busy” message while it processes the classifications. When it completes, the full set of priority values should be displayed, as shown in Figure 11-4.

Figure 11-4. Completed priority classifications generated by NUMINFER()

Numerous.ai primarily operates through formulas, but it also offers capabilities for generating, analyzing, and transforming data. These features are summarized in Table 11-4. For more detailed guidance, you can access the complete documentation by clicking the Numerous.ai icon in the Home tab of the Ribbon.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Table 11-4. The features provided by Numerous.ai

Type	Feature	Description
Formula	Generate and explain	Create new content or explain existing formulas
Data	Classify, reformat, extract, sentiment analysis	Transform and interpret data
Text	Rewrite, translate	Improve clarity or convert language
Information	Make educated guesses	Retrieve or infer general knowledge

GPT for Work

GPT for Work is another very powerful add-in that provides several functions you can use directly in the cell. GPT for Work is a tool compatible with different applications, including Microsoft Word and Google Sheets. The version specific for Excel is called GPT for Excel Word. Some of these functions include the following:

GPT_CLASSIFY( )

Takes an input text and assigns it to one category from a predefined set.

GPT_EXTRACT( )

Pulls out targeted data from text (for example, names, dates, emails, URLs, product details) and returns the results as comma-separated values.

GPT_TRANSLATE( )

Converts text from one language to another. Results improve when the source language is specified.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

TIP

For the complete list of available functions, refer to GPT for Work's official website. Use these functions just as you would the Numerous.ai functions and treat them the same way you would treat standard Excel functions.

Like other add-ins, GPT for Work is a paid plug-in. It provides free credits, and when those run out, you can purchase packs of various sizes or subscribe.

GPT for Excel Word also offers another feature, similar to Copilot. Like Copilot, GPT for Excel Word provides a chat feature where you can type specific prompts, as shown at the top of Figure 11-5, which shows the basic GPT for Excel Word interface.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 11-5. The main screen of GPT for Excel Word

Extracting New Fields from Existing Fields

To see how this feature works in practice, we'll use the dataset patients.xlsx. Open it and launch the add-in (install or start it as described earlier). In this example, we'll use

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

GPT for Excel Word to extract the city and postal code from the address field. We'll add two new columns to the table: one for the city and one for the postal code.

Follow these steps:

1. In cell K1, enter City, and in L1, enter CAP (CAP stands for Codice di avviamento postale, which is Italy's version of a postal zip code system). This expands the table by two columns, both initially empty.
2. Open the GPT for Excel Word add-in and, in the chat panel, enter the following prompt:

   ```
   Fill out the columns City and CAP from Address.
   ```

Figure 11-6 shows the resulting output, with the City and CAP columns automatically populated.

Figure 11-6. The output produced by GPT for Excel Word

The previous example demonstrated how to extract new fields from an existing one. Now we'll look at how to use GPT for Excel Word to enrich the dataset with additional information.

Enriching a Dataset with More Information

Suppose we want to add a new column that contains the patient's discharge summary. This is a prefilled, editable text that includes basic patient details and can later be expanded by the physician with more specific discharge notes.

Follow these steps:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

1. In cell M1, type Discharge Summary. The table automatically expands by adding a new column, which is initially empty.
2. In the GPT for Excel Word chat panel, enter the following prompt:

```
Generate a hospital discharge summary using the following patient data
from this row:
- Patient Name: [Name]
- Age: [Age]
- Diagnosis: [Diagnosis]
- Discharge Date: [Discharge Date]

The discharge summary should include:
- Patient demographic information
- Primary diagnosis
- Clinical status at the time of discharge (leave this section editable)
- Home treatment instructions (leave this section editable)
- Follow-up recommendations (leave this section editable)
- Signature lines for physician and patient

Use a clear, professional medical tone.
Return only the final discharge summary text.
The generated text must be placed in the column: Discharge Summary.
```

At this point, GPT for Excel Word fills the entire column with a discharge summary template. For example, in my case, the output generated for row 2 was this:

```
**Hospital Discharge Summary**

**Patient Information:**
- **Name:** Laura Jackson
- **Age:** 77
- **Discharge Date:** 25/08/2025

**Primary Diagnosis:**
- **Diagnosis:** Fracture

**Clinical Status at Discharge:**
- [Editable Section: Please provide details on the patient's clinical status at
the time of discharge, including any improvements, remaining symptoms, or
ongoing medical concerns.]

**Home Treatment Instructions:**
- [Editable Section: Please provide specific instructions for home care,
including medication regimens, activity restrictions, dietary recommendations,
and any other relevant care guidelines.]

**Follow-Up Recommendations:**
- [Editable Section: Please provide recommendations for follow-up care,
including any necessary appointments with specialists, physical therapy sessions,
or additional diagnostic tests.]
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Signatures:

_____
Physician Signature
Date: _____

_____
Patient Signature
Date: _____

This summary ensures that Laura Jackson and her healthcare providers have a clear understanding of her condition and the necessary steps for her continued recovery.

Exercise 1

Consider the patients.xlsx dataset again. Add a Length of Stay column to the table, then based on this new column, create a Severity column as follows:

- 0–3 days—severity is Mild
- 4–10 days—severity is Moderate
- 11+ days—severity is Severe

Use GPT for Excel Word to solve this exercise. You can find the solution to the exercise in 11/exercise1_solution.txt.

TIP

It's easy to feel overwhelmed by the number of available add-ins, especially if you assume each one requires a subscription or ongoing credit purchases. In practice, most add-ins offer free tiers or trial credits that let you experiment before committing. Use small sample datasets while testing so you can evaluate multiple features without quickly consuming your available credits. Once you're confident that the tool fits your workflow, then consider upgrading to a paid plan.

Power GPT

Power GPT extends Excel with intelligent assistance for working with data and formulas. It can explain how existing formulas work, generate new formulas from plain-language instructions, translate formulas across different regional syntax settings, build SQL queries based on selected ranges, and even produce regular expressions for data validation and text processing tasks.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Power GPT provides three core capabilities:

Generate

> Automatically create formulas, SQL statements, VBA snippets, regex patterns, or Python code by describing the desired outcome in natural language. This allows users to move from intent to implementation without manual syntax lookup.

Ask Excel Expert

> A conversational help interface for troubleshooting formulas, clarifying statistical functions, resolving errors, or exploring the best way to structure a calculation. It acts like an on-demand tutor integrated directly into the workbook.

Generate (formulas, SQL, VBA, regex, and Python)

> A more advanced mode that gives direct control over the type of output needed, enabling rapid prototyping of data transformations, automation scripts, and analytical workflows.

Now we'll see an example showing how to use the Ask Excel Expert feature. Open the 11/patients.xlsx file and, after installing the Power GPT add-in, launch it. From the interface, select Ask Excel Expert from the Select a GPT Function drop-down menu, as shown in Figure 11-7.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 11-7. The Select a GPT Function drop-down menu in Power GPT

Write the following prompt in the chat box:

```
From this table, filter for patients with column Outcome = 'Under Treatment',
group by Department, count patients per department, create a bar chart, and
summarize the distribution in one paragraph.
```

As an output, Power GPT provides the steps to implement our request. In my case, it provided the following answer:

```
To filter for patients with the column "Outcome" set to "Under Treatment," group
them by "Department," count the patients per department, create a bar chart, and
summarize the distribution, follow these detailed steps:
Step 1: Filter the Data
  1. Open your Excel file containing the patient data.
  2. Select the data range that includes the "Outcome" and "Department" columns.
  3. Go to the Data tab and click on Filter.
  4. Click the filter dropdown on the "Outcome" column and select only 'Under
     Treatment'.
Step 2: Group by Department and Count Patients
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

1. After filtering, you can use the PivotTable feature:
   - Go to the Insert tab and click on PivotTable.
   - Select the filtered data range and choose where to place the PivotTable.
2. In the PivotTable Field List:
   - Drag the Department field to the Rows area.
   - Drag the Patient ID (or any unique identifier) field to the Values area and set it to Count.

Step 3: Create a Bar Chart

1. With the PivotTable selected, go to the Insert tab.
2. Choose Bar Chart from the Chart options.
3. Select the type of bar chart you prefer (e.g., clustered bar).
4. Format the chart as needed (add titles, labels, etc.).

Step 4: Summarize the Distribution

After creating the bar chart, you can summarize the distribution of patients under treatment by department. Here's an example of how you might write that summary:

Summary of Patient Distribution Under Treatment

The analysis of patients currently under treatment reveals a diverse distribution across departments. The bar chart indicates that the [Department A] has the highest number of patients, suggesting a significant demand for medical resources and attention in that area. Conversely, [Department B] shows a notably lower count, which may reflect either effective patient management or fewer cases requiring treatment. Overall, this distribution highlights the varying levels of patient care needs across departments, emphasizing the importance of resource allocation to ensure optimal care for all patients.

Note:

- Replace [Department A] and [Department B] with the actual department names and their respective counts from your analysis.
- Ensure that your chart is clear and visually represents the data effectively for better understanding.

Let's follow the steps proposed by Power GPT. Figure 11-8 shows the output after applying step 1. As shown in the figure, I have used a filter in the Output column (indicated by the funnel icon).

Figure 11-8. The result after following step 1 proposed by Power GPT

Figure 11-9 shows the result after following steps 2 and 3. On the left, there is a PivotTable with the filtered values, and on the right, the corresponding bar chart.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 11-9. The result after following steps 2 and 3 proposed by Power GPT

THE INVISIBLE ORCHESTRA

I've loved music for as long as I can remember. Even now, while I'm working, I often have music playing in the background. It brings a touch of creativity into my day, like a gentle spark that opens up space for new ideas. When I listen, I sometimes think about how much effort goes into creating a single song. There's the person who writes the lyrics, the one who shapes the melody, the musicians who bring the sound to life, and the producer who refines it all.

By the time the song reaches us, all we hear is the finished piece. We don't see the layers of work behind it. And yet, it moves us. It reaches us in ways we can't always explain. Music fills my heart and reminds me that creation is always a collective act.

AI tools are similar in that way. Each one contributes just a part of the workflow. One extracts data, another rewrites text, another helps automate, and another explains. On their own, they may seem small. But together, when orchestrated with intention, they create something cohesive, powerful, and sometimes even beautiful. They allow you to shape something larger than the sum of its parts.

The art is not in any single tool. It's in how you bring them together.

Sally Suite

Sally Suite is an Excel add-in that provides functionality comparable to Copilot. Once installed, it adds its own tab to the Ribbon (Figure 11-10), giving you direct access to a

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

range of AI-assisted tools. These include data analysis features and the ability to generate VBA macros directly from natural language instructions, streamlining both exploratory work and automation tasks.

Figure 11-10. The Sally Suite tab in the Ribbon

Let's focus on the Data Analysis feature. Open the patients.xlsx dataset. Suppose you want to compute the number of patients in each department. First, open the Sally Suite tab from the Ribbon, then click Data Analysis. This opens a side panel. In the lower-left corner, you can choose which AI model to use. At the time of writing, available models include GPT-4o mini (default), Gemini, DeepSeek, and others. Keep the default model selected, and in the chat box enter the following prompt:

```
Consider Table1. Calculate the number of patients for each department.
```

The assistant processes the request and then asks you to confirm the results it intends to produce. Click Confirm. The output is generated directly in the worksheet and closely mirrors the style and structure of a Copilot-generated result, as shown in Figure 11-11.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 11-11. The table generated by the Sally Suite assistant

Click New Sheet to insert the resulting table into a separate worksheet. Note that the values are returned as static results rather than as dynamic formulas.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

A particularly noteworthy feature of Sally Suite is that it also provides the corresponding Python code used to compute the results. If you are familiar with Python, you can copy this code directly from the chat panel, paste it into your own script file, and run it independently. Figure 11-12 shows the section of the chat interface where the Python code can be copied using the Copy button.

Figure 11-12. The buttons to access the Python script in Sally Suite

For reference (or simply out of curiosity), the Python code generated for this operation is shown next. If you're not comfortable with programming, feel free to skip this section:

```
import pandas as pd

def main():
    # Read the data from the specified Excel file and sheet
    file_path = '/input/data.xlsx'
    try:
        # Loading the patient data from the Excel sheet
        df = pd.read_excel(file_path, sheet_name='patients')
    except Exception as e:
        print(f'Error reading Excel file: {e}')
        return
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
    # Validate that the relevant columns are in the dataframe
    if 'Department' not in df.columns:
        print('Department column is missing from the data.');
        return

    # Count the number of patients in each department
    department_counts = df['Department'].value_counts().reset_index()
    department_counts.columns = ['Department', 'Number of Patients']

    # Write the results to a Markdown file
    output_file = '/output/patient_department_counts.md'
    with open(output_file, 'w') as file:
        file.write('# Patient Count by Department\n\n')
        file.write('| Department | Number of Patients |\n')
        file.write('|------------|--------------------|\n')
        for index, row in department_counts.iterrows():
            file.write(f'| {row.Department} | {row["Number of Patients"]} |\n')

    print(f'Results saved to {output_file}')

main()
```

Exercise 2

Open the patients.xlsx dataset. Use Sally Suite to write a macro that filters Table1 to show only rows where Outcome = 'Discharged' and then formats the resulting rows with a yellow background. The solution to this exercise is available in 11/exercise2_solution.xlsx.

Summary

In this chapter, you explored a snapshot of the AI add-ins currently available in the Excel marketplace. This is only a starting point: new tools appear regularly, and I encourage you to browse the Microsoft Office Add-in Marketplace to find add-ins that fit your workflow and data needs.

Here are the takeaways from this chapter:

- AI add-ins extend Excel with specialized capabilities. They allow users to embed AI directly into formulas, automate workflows, and perform structured, repeatable data transformations.
- AI add-ins differ from Copilot in scope and usage. Add-ins are most effective for precise, repeatable spreadsheet tasks inside Excel, while Copilot is

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

designed to support broader, cross-application workflows across Microsoft 365.

AI in Excel can play four main roles: as a formula that processes data cell-by-cell, as automation that executes end-to-end workflows, as an explanation tool that clarifies logic, and as documentation that generates professional text.

Numerous.ai provides formula-based AI functions. It allows the user to describe a transformation in plain language and apply it directly to cells, enabling tasks such as data extraction, example-based classification, and rewriting existing content.

GPT for Excel Word combines in-cell AI functions with a chat interface. It is particularly effective for extracting structured information from text, translating data, or generating output text such as summaries or report templates directly into spreadsheet cells.

Power GPT focuses on generating formulas, SQL queries, Python scripts, regular expressions, and VBA code from natural language instructions. It also provides a conversational assistant to explain formulas and guide problem-solving.

Sally Suite offers AI-assisted analytics and workflow support inside a dedicated Excel Ribbon tab. It can analyze tables, generate summary tables and visualizations, and provide the corresponding Python code used to produce the output.

In the following chapter, we'll shift focus from add-ins to Python. We'll look at how to use Python alongside Excel and AI to build more advanced, programmable workflows that go beyond what the add-ins alone can provide.

Activity

Use the Caesar cipher (shift each letter seven positions backward in the alphabet) to decipher the following sentence, which summarizes the chapter you have just read:

> HP hkk-puz luohujl Lejls if vmmlypun zwljphspglk avvsz mvy khah huhsfzpz huk hbavthapvu.

The solution to this activity is in Appendix D.

Chapter 12. Python, Excel, and AI

I'm a computer engineer, and programming has been a steady companion for many years, so writing a chapter that brings together Python, AI, and Excel is something I genuinely enjoy. It wasn't always that way, though. I still remember my first encounter with a programming language more than 20 years ago—it was far from smooth. In my very first computer science course, we learned C++. The challenge wasn't just the language itself but the shift in mindset it required. Programming forces you to stop thinking from your own perspective and start thinking from the machine's. The key question becomes: how would a computer implement this? Once that mental switch happens, everything feels more natural because you begin reasoning the way the system does.

In this chapter, I want to give you a taste of what's possible when you extend Excel with Python, one of the most popular programming languages for data analysis. Using Python turns Excel into an even more potent ally. You don't need to become a full-fledged programmer either. Today, AI can support you even if your coding experience is limited.

The real point of this chapter is simple: Python gives you a flexible, unifying scripting layer that connects data access, analysis, and visualization in a way that feels natural for AI-driven workflows. You'll need to spend a little time getting Python set up in Excel, but that upfront investment pays off quickly in power, clarity, and control.

In this chapter, you'll explore these topics:

- Embedded Python
- Architecture using Excel, Python, and the OpenAI API
- Environment setup
- Using Python and OpenAI for sentiment analysis
- Using AI tools, Python, and OpenAI for classification

Let's start with the embedded version of Python.

Embedded Python

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Before we start to integrate external APIs such as OpenAI in Excel, you will benefit from knowing what Python in Excel can do. Microsoft provides a built-in Python environment for Excel. This feature is convenient for basic analysis and quick prototypes, but it has strict limits.

Python in Excel is a Microsoft 365 feature (not available on iPad, iPhone, and Android devices) that evaluates Python code inside a worksheet. This feature uses the =PY( ) function. Consider something like this:

```
=PY("sum([1, 2, 3])")
```

Excel sends this code to a secure cloud environment. The environment returns the result to your workbook as a number, a table, or a chart. Here, the code in brackets asks Python to sum the numbers 1, 2, and 3. The environment includes popular libraries such as Pandas for data analysis and Matplotlib for charts. You can use these tools to perform quick exploration, cleaning, or visualization tasks without leaving Excel or installing additional packages.

The Dataset

We will focus on the travel dataset available in the book's GitHub repository at 12/travels.xlsx; it contains 10 fictional entries describing recent trips to cities around the world. Each entry captures the essential details of a journey and includes the following fields:

City

: The destination you visited

Country

: The country where the city is located

Days Stayed:

: The number of days spent in that location

Avg Daily Spend (€)

: The approximate average amount spent per day during the trip

Trip Type:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

A short label that describes the nature of the trip, such as cultural, adventure, or food-oriented

Satisfaction (1–5)

A personal rating of the overall experience, with 1 being poor and 5 being excellent

Review (EN)

A brief free-text evaluation of the visit, written in English, expressing positive, mixed, or negative opinions

Figure 12-1 shows the travel dataset.

Figure 12-1. The travel dataset

Testing the Python Environment

Just for a quick test of the PY( ) function, open the 12/travels.xlsx dataset, select cell C13, and access the embedded Python editor from the Formulas tab of the Ribbon (Figure 12-2).

Figure 12-2. Python commands in the Formulas tab of the Ribbon

A new pane appears on the right. Click the “Add Python cell” button in Travels!C13. The Python editor should now appear, as shown in Figure 12-3.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 12-3. The embedded Python editor in Excel

Now you have a Python environment in cell C13. For example, to calculate the average number of days stayed in a city, write the following Python code in the Python editor:

```
days=xl("C1:C11", headers=True)
days['Days Stayed'].mean()
```

This code uses the built-in `xl()` function, which enables you to access Excel content in Python. The function receives the cell range as the first parameter, along with optional parameters. For example, setting `headers=True` treats the first passed row as the column name. The function returns a Pandas DataFrame, which can be used in Python. If you are not familiar with the Pandas DataFrame, a DataFrame is like a table where you access columns by name. For example, row 2 in the previous code snippet selects the column `Days Stayed` in the DataFrame and calculates the mean. Then, it returns the results as the cell output. Click the Save button in the Python editor to run the code. Now, cell C13 should contain the average value (4.4). In some cases, there may be a slight delay before the result is visible.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

TIP

Alternatively, you can write the Python code directly in the selected cell. Type =PY(in cell C13 to enable the Python environment. As soon as you write that (yes, without closing brackets), the cell format changes, as shown in Figure 12-4.

Figure 12-4. The embedded Python environment enabled within a cell

TIP

As suggested in the Help sidebar, you can exit the Python environment and run the code in the cell by pressing Control (or Command on Mac) + Return.

The previous example is very simple, and, usually, you don't use Python to perform an operation that you can do in a more straightforward way using the Excel direct functions. However, this example illustrated how the embedded Python version works in Excel.

WARNING

The embedded Python environment has strict limits. For example, the runtime does not allow internet access, which prevents any calls to external services. For this reason, you can't integrate OpenAI or any other AI model directly inside =PY(). The environment also does not allow installing new Python libraries, so you must rely on the ones Microsoft preinstalls.

Using Generative AI Tools

Although the embedded Python environment does not permit the direct integration of external AI libraries, you can still use your preferred AI tool, such as Copilot or ChatGPT, as an assistant to write Python code. As I've often said throughout this book, I suggest using the story structure to build your prompt: context, main point, action.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

To understand how this approach works, let's revisit the Travels dataset. We will use Copilot as an AI tool (although you can also use other AI tools, such as ChatGPT). We will use Copilot to write the Python code to produce a chart showing the total spending for each city in the dataset, sorted in descending order.

Open the 12/travels.xlsx dataset and launch Copilot from the Home tab of the Ribbon. Write the following prompt in the Copilot chat:

```
Use the PY() function provided by Excel. Consider cells A1:G11, write the code in
Python that:
  1. Selects all the table cells using xl() with headers=True and stores them in
     a variable called df
  2. Calculates the total spending for each city as the product between the Avg
     Daily Spend and Days Stayed columns and stores it as a new column of df
  3. Orders df by decreasing total spending
  4. Draws a bar chart using Matplotlib showing the total spending vs. the city
```

Table 12-1 shows how each part of the prompt maps to the story structure.

Table 12-1. Mapping between the prompt and the story structure

Story structure	Part of the prompt
Context	Use the PY() function provided by Excel. Consider cells A1:G11, write the code in Python that:
Main point	Selects all the table cells using xl() with headers=True and stores them in a variable called df Calculates the total spending for each city as the product between the Avg Daily Spend and Days Stayed columns and stores it as a new column of df Orders df by decreasing total spending
Action	Draws a bar chart using Matplotlib showing the total spending vs. the city

In my case, Copilot provided the following Python code:

```
import matplotlib.pyplot as plt

# Select the table with headers
df = xl("A1:G11", headers=True)
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
# Calculate total spending
df["Total Spending"] = df["Days Stayed"] * df["Column1"]

# Sort by total spending descending
df = df.sort_values("Total Spending", ascending=False)

# Plot bar chart
plt.figure(figsize=(8, 4))
plt.bar(df["City"], df["Total Spending"], color="skyblue")
plt.xlabel("City")
plt.ylabel("Total Spending")
plt.title("Total Spending by City")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Review the generated code and, if necessary, correct any cell ranges or other issues. In my case, the generated code was correct. Copy the generated code and then do the following:

1. Select cell D16 and start writing the formula =PY((without the closing bracket). As shown in Figure 12-4, the Python environment is now enabled.
2. Paste the copied Python code generated by Copilot.
3. Press Control (or Command on a Mac) + Return. Cell D16 should now show the value [PY] Image.
4. Click cell D16 and the small table icon, named Insert Data, on the top right of the cell.
5. Select Display Plot over Cells from the drop-down menu, as shown in Figure 12-5.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 12-5. The drop-down menu for a cell containing a Python formula

Now you should see the final chart, as shown in Figure 12-6.

Figure 12-6. The final bar chart generated using Python and Copilot

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

TIP

If your Python code contains an error, the cell with the error will show the value #PYTHON. To correct the error, right-click the cell and select Show Editor from the drop-down menu. The editor appears on the right, showing the error type as well. Correct the error and click the Save button to run the code. You can even ask Copilot to correct the error.

Exercise 1

Consider the 12/films.xlsx dataset available in the book's GitHub repository. Using the embedded Python environment, calculate the length of each review and add it as a new column. The solution to the exercise is available at 12/exercise1_solution.txt.

Architecture of Python Add-ins

The embedded version of Python works well for small analytical tasks, but as mentioned it cannot reach the internet or load libraries that aren't already preinstalled. Many real-world projects require access to advanced packages, API calls, or custom automation. Python add-ins fill this gap. They connect Excel to the Python installation on your computer and allow two-way communication between the worksheet and your scripts.

With an add-in, Excel can send cell values to Python, run your code, and receive processed results. This approach turns Excel into a flexible frontend and lets Python act as the primary execution engine.

In this chapter, we focus on the xlwings add-in for running Python in Excel. xlwings is a lightweight, beginner-friendly tool. It lets you run Python functions from Excel buttons or macros. It can read and write cells through a simple API. This makes it suitable for tasks such as sentiment analysis, summarization, or any automation that requires sending text to an AI model and returning the result to the worksheet.

When you use xlwings, you are working inside a small, distributed system with several components that cooperate to move data, run code, and return results. Understanding this architecture helps you design workflows that make sense and avoid common mistakes.

Although every add-in has its own implementation details, their overall structure is similar. This chapter focuses specifically on xlwings as an add-in and the OpenAI API as an external API. The system has four main pieces, shown in Figure 12-7.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 12-7. The architecture of the system integrating Excel, Python, and OpenAI

Each plays a specific role, and together they form a loop of data and control.

Excel

Excel remains the frontend. It stores the dataset, receives outputs, and provides the user with the interaction surface. Excel does not run Python on its own. Instead, it communicates with the add-in whenever a user clicks a button, triggers a macro, or enters a UDF (user-defined function).

Excel provides the following:

- Cell ranges that become inputs for Python
- Macros or formula calls that trigger code execution
- A place to display results such as sentiment labels, summaries, charts, and categories

The add-in is the most critical component. It acts as the mediator that connects the spreadsheet to your Python environment. Without an add-in, Excel has no way to call Python, share data with it, or receive values back.

xlwings

The xlwings add-in performs three essential duties:

Launches Python code

When you press a button or call a formula, the add-in loads your Python script and calls the matching function.

Moves data between Excel and Python

The add-in translates cell ranges into Python data structures.

Returns results to Excel

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

After Python finishes processing, the add-in writes the result back to the worksheet or returns it to the formula caller.

Python

Once the add-in hands off the request, your local Python installation performs the computation. This part of the system gives you complete control, because it uses the same Python you have on your machine.

The components involved here include the following:

- Your script files
- Installed Python packages, such as Pandas, OpenAI, or NumPy
- Your OpenAI client configuration
- Your environment variables, where the API keys are stored

From Excel's point of view, Python is a black box. Excel sends a request, waits, and receives a result. Python handles everything in between.

OpenAI API

When your Python script calls the OpenAI API, a fourth component comes into play: the external AI service. The sequence looks like this:

1. Excel sends data to Python (through the add-in).
2. Python formats the request.
3. Python sends the request to OpenAI over the internet.
4. OpenAI runs the model and returns the response.
5. Python processes the response and returns it to Excel through the add-in.

OpenAI is not aware that Excel is involved. From its point of view, a Python program simply sends a request.

Environment Setup

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Now that you have learned the architecture behind the combination of Excel, Python, and AI, you are ready to set up the environment. Let's start with installing Python on your computer.

TIP

If you are using Windows, download and install PowerShell, a powerful terminal with many features comparable to those in Linux and macOS. More details about the installation are available from Microsoft.

Installing Python on Your Computer

The simplest way to install Python is to use the standard installer from python.org. Follow these steps:

1. Visit python.org and download the latest stable release for your operating system.
2. Run the installer and enable the option that adds Python to your system path.
3. Open a terminal and verify the installation with this command:

```
python --version
```

Or this one, depending on your OS:

```
python3 --version
```

You should see something like this:

```
Python 3.11.3
```

The Python version changes based on what you have installed.

After the installation works, install the essential packages:

```
pip install pandas xlwings openai
```

Or, depending on your OS:

```
pip3 install pandas xlwings openai
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

If the installation is successful, you should see a message in the terminal containing the following text:

```
Successfully installed [...]
```

In case of an installation issue, you can describe the problem and ask an AI tool to help you resolve it.

Configuring Access to the OpenAI API

OpenAI requires an API key and credit for use. The key identifies your account and grants permission to use the models. You create the key in your OpenAI dashboard:

1. Visit OpenAI Platform and register a new account if you don't have one. Then log in.
2. Click Quickstart → Create an API Key and follow the guided prompts. Copy and paste the API key into a file called openai_key.txt.
3. Click the Settings icon near your avatar at the top right of the screen. Then select Billing from the left menu. Follow the guided prompts to add credit. You need to configure your credit/debit card. A credit of $5 is sufficient for our experiments. Billing is per use rather than on a subscription, so if you have to repeat calls because you made a mistake or start experimenting, you may be charged.

The next step is to check if everything works. If you are using Windows, download the GitHub repository for the book to your local filesystem. Then, go to the directory on your filesystem that contains the book's GitHub repository and navigate to the subdirectory 12 for chapter 12, right-click anywhere in the directory, and select Open in Terminal from the drop-down menu. In Linux/macOS environments, open a new terminal and move to the 12 directory through the cd (change directory) command.

We will run a simple test to verify that the environment is set up correctly using the following code, which you can also find in the GitHub repository for the book at 12/test.py:

```
from openai import OpenAI
import os

def load_api_key(path="openai_key.txt"):
    with open(path, "r", encoding="utf-8") as f:
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
        return f.read().strip()

client = OpenAI(api_key=load_api_key())

response = client.responses.create(
    model="gpt-4.1-mini",
    input="Test message"
)

print(response.output[0].content[0].text)
```

After importing the required libraries with the `import` command, the script defines a function that loads the OpenAI API key from the openai_key.txt file (following the unique path for that file on your computer, which you entered). Next, it creates an OpenAI client that uses the loaded key and sends a simple test message to the OpenAI service using the `create()` function. If the connection works properly, the OpenAI service will respond with a text message greeting you or something similar.

To run the Python script, type the following command in the terminal:

```
python test.py
```

Or, depending on your OS:

```
python3 test.py
```

If the test is successful, you should see something similar to the following output:

```
Hello! How can I assist you today?
```

If you don't see such text, there is a problem with the OPENAI_API_KEY configuration, so I suggest you paste the error into your AI assistant to search for a solution.

Installing the xlwings Add-ins in Excel

xlwings comes in two flavors:

- xlwings Standard runs Python installed on your machine.
- xlwings Lite is a web-based add-in that runs Python in the browser via Pyodide.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

We'll use xlwings Standard because it gives us full access to the Python ecosystem and allows code to interact with the local filesystem. To install it, run the following command in the terminal:

```
pip install xlwings
```

Or, depending on your OS:

```
pip3 install xlwings
```

Then register the Excel add-in by running the following command in the terminal:

```
xlwings addin install
```

Once installed, xlwings adds a tab in the Excel Ribbon that lets you run Python directly from your workbooks. After running the previous command, open Excel and check if this new tab appears in the Ribbon. You should see something similar to the tab shown in Figure 12-8 (Windows) and Figure 12-9 (macOS).

Figure 12-8. The xlwings tab in the Ribbon (Windows)

Figure 12-9. The xlwings tab in the Ribbon (macOS)

As you can see from the figures, the Windows version provides additional features, including support for user-defined functions (UDFs). This feature is also available in macOS with the PRO version of xlwings.

The next step is to ensure the Python interpreter is configured correctly. If the text box near the interpreter is empty, you must add the path to the Python interpreter manually. On Windows, open the Start Menu, search for Python, right-click the entry, and choose "Open file location" to reveal the actual executable. Then right-click anywhere in the directory and select Properties from the drop-down menu. Copy the path corresponding to the Location label and paste it into the Interpreter text box on the xlwings tab.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

On macOS and Linux, run the following command in the terminal:

```
which python3
```

This command prints the location of the active Python interpreter.

THE JOY OF FINALLY MAKING IT WORK

At this point, the entire system is configured and ready to use. But before we move on, let me share a small personal story. When I was in university, I had to install a whole web server for a web design and programming course. This was about 20 years ago, and things were much more complicated back then. You had to install everything manually, just as you've done here.

Today, setting up a web server is trivial: you download a bundled package and follow an installer wizard. Back then, nothing was that simple. You could spend days trying to get all the components in place, often getting frustrated because something went wrong and you had no idea who to ask. There was no AI to paste your error into and get an answer within seconds. Instead, you asked a classmate or a friend—but never the professor, who absolutely did not want to be bothered with installation issues. So you struggled for days, chasing down obscure problems, until that one classmate, someone you barely knew but who happened to understand the issue, showed up and solved it in a matter of minutes.

The relief and joy were enormous. I still remember the day I finally installed Apache and deployed my first web application. It felt like a real triumph. I hope you feel a bit of that same satisfaction now that you've installed all the components needed to bring Python to Excel.

Sentiment Analysis

Sentiment analysis is a quick way to gauge how someone feels without reading everything in detail. It's perfect for reviews, feedback, and surveys when you just want to take the emotional pulse.

Writing the Script

We will use the 12/sentiment.py script available in the book's GitHub repository. This script calculates the sentiment for each review in the travels.xlsm dataset and writes the

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

result into a new column.

Before running the script, let's analyze what it actually does. First, it imports the required libraries and loads the OpenAI API key from the openai_key.txt file, as shown in the following code:

```
import os
import xlwings as xw
from openai import OpenAI

def load_api_key(path="openai_key.txt"):
    base = Path(__file__).parent
    abs_path = (base / path).expanduser().resolve()
    with abs_path.open("r", encoding="utf-8") as f:
        return f.read().strip()

os.environ["OPENAI_API_KEY"] = load_api_key()

client = OpenAI()
```

Next, the script defines a function named main(), which connects to the Excel workbook, selects the first worksheet, and adds a new column header, as shown in the following code:

```
def main():
    wb = xw.Book.caller()
    sheet = wb.sheets[0]
    sheet["H1"].value = "Sentiment"
```

Now the script iterates over all the rows (except the first, which contains the header) and extracts the review from column G:

```
    row = 2
    while True:
        review = sheet[f"G{row}"].value
        if review is None:
            break
```

Continuing in the loop across the rows, the script defines a prompt to extract the sentiment for the current review (passed as an argument):

```
        prompt = (
            "Classify the sentiment of this travel review as "
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
        "'positive', 'negative', or 'mixed'. "
        "Return only the single word label.\n\n"
        f"Review: {review}"
    )
```

It passes the prompt to the GPT-4.1 mini model and writes the answer (stored in the `response` variable) in column H of the worksheet:

```
    response = client.responses.create(
        model="gpt-4.1-mini",
        input=prompt
    )

    label = response.output[0].content[0].text.strip()
    sheet[f"H{row}"].value = label

    row += 1
```

Finally, the script defines the code to run it from the command line:

```
if __name__ == "__main__":
    xw.Book("travels.xlsm").set_mock_caller()
    main()
```

Running the Script

There are two ways to run the script. The first uses the command line. Open the terminal and use the `cd` command to navigate to the directory containing your Python script, following the steps described previously. Then write the following command:

```
python sentiment.py
```

Or, in macOS:

```
python3 sentiment.py
```

The first time you run this script, a dialog box may open asking you to grant the script access. Accept and proceed.

If the script is successful, you won't see any output from the command line. Open the travels.xlsm file. You should see a new column added to the dataset, as shown in Figure 12-10.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 12-10. The new column added after running the script

The second way to run your Python script is directly from Excel. In this case, you must set up a macro that runs the Python script. Ensure that your workbook has macros enabled. To run your script from Excel, perform the following steps:

1. Access the xlwings tab from the Ribbon and select the option "Add workbook to PYTHONPATH." This ensures that Excel can read scripts located in the same directory as the Excel file.

2. Open the Developer tab on the Ribbon and click Macros. In the text box under "Macro name", type `sentiment` (the name for the macro) and click the Create button.

3. Write the following VBA code in the VBA Editor:

```
Sub sentiment()
    Application.Run "xlwings.RunPython", "import sentiment; _
    sentiment.main()"
End Sub
```

4. This code invokes the `RunPython` command from the xlwings library and executes the `main()` function in the sentiment file. Note that the code does not include the .py extension.

5. Click the Save button and then click Run (the triangular "Play" button in the VBA Editor's toolbar). The script should run correctly, and you should see the output shown in Figure 12-10.

Classification

You are approaching the finish line, and this is the last example in the book. To wrap things up with a bit of flair, we'll bring together several concepts introduced earlier, some of which appeared on their own. I hope this final example gives you the boost you need to see how powerful AI can really be.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

We're about to build a complete workflow: classifying the destinations in travels.xlsm into three categories: highly recommended, recommended, and not recommended. We'll generate the Python function that performs the classification and then run the process as described in the previous section.

Generating the Python Code with AI

Here's the twist: this time, we will not write the function ourselves. Instead, we will ask an AI assistant, such as ChatGPT or Copilot, to write the code for us, then review and fine-tune it. The real skill, as always, lies in writing a good prompt that breaks down into the three parts you should be familiar with. To make AI write a correct Python script, the prompt needs to define these three parts:

Context: Which dataset to use, ideally attached in the chat

Main point: The detailed behavior that the Python code must implement

Action: What the code should return

Here is what such a prompt might look like for the Travels dataset:

```
Consider the attached dataset. Write a Python function using xlwings and OpenAI
that classifies each row in the following categories: highly recommended,
recommended, not recommended. The classification must be based on the review
text and any other relevant fields in the dataset. The function must read the
values from the worksheet using xlwings, apply the model to each row, and write
the resulting category back into Excel. Call the main function main(). When you
use OpenAI import the OPENAI_API_KEY using this code:
from pathlib import Path def load_api_key(path="openai_key.txt"):
    base = Path(__file__).parent
    abs_path = (base / path).expanduser().resolve()
    with abs_path.open("r", encoding="utf-8") as f:
        return f.read().strip()
os.environ["OPENAI_API_KEY"] = load_api_key()
client = OpenAI()
```

To make AI generate the code for you, perform the following steps:

1. Access your preferred AI tool and write the preceding prompt.
2. Save the generated code in a file named 12/classify.py.
3. Ensure the code is correct (at least in general).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

4. Run the code using the procedure described in the "Sentiment Analysis" section ("Running the Script").

Using AI for Troubleshooting

In my case, I obtained the output contained in 12/classification_wrong.py. If you try to run my code, it produces a run-time error. This should remind you that the code generated by AI is not perfect, but you can use the same AI to correct the mistakes it made.

Follow these steps to use AI to correct the errors:

1. Copy the error to the same chat where you generated the code.
2. Your AI assistant generates a corrected version of your code or part of your code.
3. Correct the original code provided by the AI assistant.
4. Ensure the code is correct (at least in general).
5. Run the code.
6. Repeat steps 1–5 if necessary.

You can find the correct version of the code in 12/classification.py. Run the code using the steps described previously in the "Sentiment Analysis" section ("Running the Script"). Figure 12-11 shows the final result.

Figure 12-11. The Travels dataset after running the classification script

Exercise 2

Consider the 12/films.xlsx dataset available in the book's GitHub repository. Write the Python code that will read the movie review dataset directly from Excel using xlwings and send each review to an OpenAI model to extract its sentiment label. The code must iterate over all rows, classify each review as positive, negative, or mixed, and write the

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

resulting sentiment back to a new column in the worksheet. The goal is to reproduce the workflow introduced earlier in the chapter and apply it to a new domain: film reviews. The solution to the exercise is available at 12/exercise2_solution.py.

Summary

In this chapter, you've seen some foundational techniques for bringing Excel, Python, and AI together into a single, practical workflow. You explored how Python can extend Excel beyond traditional formulas, how data can move more fluidly between analysis and automation, and how AI-powered tools can add a layer of intelligence on top of your existing spreadsheets. Here are the key takeaways:

- Embedded Python in Excel lets you run small Python analyses directly in worksheet cells, enabling quick experimentation in a controlled environment without internet access.
- An architecture based on Excel, Python, and the OpenAI API clearly separates responsibilities: Excel as the interface, Python as the processing layer, and OpenAI as the external AI service.
- You can use Python and the OpenAI API in Excel to implement different tasks, including sentiment analysis and classification.

Activity

Solve the cryptogram in Figure 12-12 using the numbers already decoded in the keyword. Use the known values to help decode the remaining letters. You can find the solution in Appendix D.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure 12-12. Cryptogram to solve

Final Thoughts

We've reached the end of our journey together, and I have to admit it feels a bit emotional. You started with your first steps in AI and Excel and arrived at full automation with Python. Technology moves fast, but what I've truly hoped to give you was not just rules or syntax. I wanted to share a way of thinking that you can carry into any new scenario you face.

I've really enjoyed walking this path with you. Even if we've never met, it still felt like a real connection took shape, and saying goodbye now is not easy. If we meet in the future, I'll be genuinely happy. Until then, I ask only one thing: do not stop learning. Keep reading, keep exploring, keep improving. There is so much knowledge out there waiting for you.

Throughout the book, I also shared pieces of my own life because learning never stands apart from living. And since you already know how much I love stories, here is one last thought. Stories usually have three objectives: they inform, they help us make decisions, and they entertain. This book was built exactly on those three levels. The theory and examples are meant to inform. The exercises invite you to think and decide. The activities and the applied fiction in the appendices are there to entertain. Indeed, there was a fourth objective too, one that mattered just as much as the others: awakening emotions. Through the personal stories I shared, I hope to do more than teach or entertain. I hope to reach you on a deeper level, to touch something human and warm

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

inside you. Emotions are bridges. They exist to connect one heart to another, to create a moment where two people who have never met can still understand each other. That was my intention from the very beginning, and I truly hope I managed to do that with you.

If this journey taught you something, inspired you, or simply kept you company, then every moment spent writing it was worth it.

Top Courses at No Cost:

- **Adobe Illustrator Course**
- **Adobe InDesign Course**
- **Adobe Lightroom Course**
- **Adobe Photoshop Course**
- **Adobe Premiere Pro Course**
- **Affiliate Marketing Course**
- **AI SEO Course**
- **Data Science Full Course**
- **English Speaking Course**
- **Python Course**

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Part IV. Stories and Solutions

Part IV is organized as a hands-on companion to the first three parts of the book, bringing together a collection of practical activities designed to reinforce the foundational skills introduced in the early chapters. The appendices focus on applied exercises that help you consolidate what you've learned by working directly with realistic spreadsheet scenarios. Part IV is structured as a short novel that defines what I call applied fiction, a technique for immersive learning.

Appendix A presents hands-on activities that reinforce the key topics from Part I of the book, using ChatGPT as a practical support tool. You'll begin with data organization and visualization, practicing how to structure data clearly and create charts that communicate insights effectively. Next, you'll work on building reusable templates, learning how to standardize layouts, formulas, and formatting to make recurring tasks faster and more consistent. Finally, you'll explore basic macro creation, automating repetitive actions with simple recorded and edited macros to improve the efficiency of your workflow.

Appendix B is organized as a practical extension of Part II of the book, offering hands-on activities that help you apply Copilot to real data analysis workflows. The exercises are structured to follow the natural progression of a data project: you begin by preparing the data with Copilot, cleaning and structuring datasets for analysis. You then move on to basic statistics, using Copilot to calculate and interpret key measures. Next, the focus shifts to data exploration and visualization, where you uncover patterns and trends through charts and summaries. Finally, the appendix concludes with data reporting, guiding you in organizing results into clear, well-structured outputs suitable for sharing with others.

Appendix C provides hands-on activities that extend Part III of the book, combining AI tools and Python in Excel. You'll start with data cleaning using Numerous.ai, and then move to data extraction with the `PY()` function to process and reshape data using Python. The appendix concludes with the use of Python and AI for prediction, introducing simple modeling and forecasting tasks that demonstrate how AI-driven code can generate forward-looking insights.

Appendix A. Before the Nepal Peaks

This appendix is the first in a three-part series where I introduce a teaching approach I call applied fiction, a fresh way to learn through storytelling. In applied fiction, the narrative isn't just for entertainment; it's a vehicle for teaching real concepts and skills. As you follow the story, you'll naturally absorb the how-to knowledge woven into it, learning by experiencing rather than by rote instruction.

The short novel unfolds across three appendices, with each part paired with hands-on activities tied directly to the book's concepts. You can read all three for the whole arc, or dip into just one and still get a complete set of activities—perfect if you want to focus on skill-building without committing to the entire story.

This appendix reviews concepts from Part I, "Using ChatGPT in Excel".

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Prologue

My name is Helena. I'm 32, and the mountains have always lived inside me, quiet, immense, irresistible. I don't remember choosing them; they simply claimed me. Over the past two years, I've climbed six mountains, standing where the air thins and every breath feels earned, where the world below shrinks into something distant and unreal.

My brother Mark understands that feeling better than anyone. The mountains are our shared language. We climb together whenever life allows it, trusting each other with ropes, with weather, with silence. Three years ago, he finally touched the dream he'd been carrying since we were kids: he landed a job in Nepal as an authorized guide for the highest peaks on Earth, the Himalayas. When he told me, his eyes were shining. I thought I'd never seen him happier.

Then, two months ago, he came home. Something in him was different. His smile didn't reach his eyes; his restlessness filled the room. One evening, as dusk bled through the windows, he finally spoke.

"Helena," he said, lowering his voice, "something strange is happening. But I can't tell you yet."

A chill crept under my skin.

"Be careful, Mark," I told him. I tried to sound calm.

"I will be."

Before leaving again, he did something that made no sense at the time. He didn't take his laptop with him. Instead, he placed it in my hands. It felt heavier than it should have.

"Take care of it," he said. "It might be useful."

Then he was gone. Since that day, I've spoken to him only twice on brief calls, rushed, his voice thin and distant, as if carried by the wind.

Today, I'm sorting through photos from my last climb, reliving moments of cold sun and aching legs. My phone rings. My heart stutters. Mark, I think, is already reaching for me.

"Helena."

It's not him.

"Oh, John," I say, unable to hide my disappointment. John is a friend of mine.

"I need to tell you something," he says. "Sit down."

The room seems to tilt. "Okay. Is something wrong?"

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

“Are you sitting?”

“Yes.”

“I read in local Nepalese newspapers that Mark has disappeared. No one has seen him for several days.”

The words don’t land at first. They hover, meaningless.

“He may have vanished during one of the climbs. I’ll keep you updated. I promise.”

The call ends. The silence roars. Dread floods me, thick and suffocating. Mark’s worried eyes, his half-spoken warning, it all crashes back at once. My legs give out. I sob until my chest burns, until my hands won’t stop shaking. Hours pass and time dissolves. Finally, I wipe my face and stand. I can’t sit with this helplessness. My gaze drifts to the laptop resting where he left it.

“I’ll turn it on,” I whisper, as if the house is listening. “He said it might be useful.”

I open the lid. The computer groans, flickers, struggles awake. The fan whirs with labored breath. Then the desktop appears. One folder sits there alone: 8848 - Not Yet. The number slams into me. Eight thousand eight hundred forty-eight meters. Everest.

Activity 1: Getting Started

The folder is a mess, containing Excel files, half-finished plots, and garbled logs. The first file that draws my attention is the map shown in Figure A-1.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure A-1. The Everest map showing Base Camp and the other four camps

I immediately recognize Mark's handwriting. The map shows the path to the Everest peak. Starting from the bottom of the map, there is the Base Camp at 5,364 m (17,598.43 ft), followed by Camp 1 at 6,065 m (19,898.29 ft), Camp 2 at 6,400 m (20,997.38 ft), Camp 3 at 7,162 m (23,497.38 ft), and Camp 4, at 8,000 m (26,246.72 ft). Finally, there is the peak at 8,848 m (29,028.87 ft).

There is another file that draws my attention. It's a worksheet named AppendixA/health_log.xlsx. I open it in Excel. It contains the data shown in Figure A-2.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure A-2. The health_log.xlsx dataset found on Mark's computer

Another file in the directory describes the dataset fields:

Altitude_m

Altitude in meters above sea level at which measurements were taken during the Everest ascent

Person_A_HeartRate

Heart rate (beats per minute) of Person A at the corresponding altitude

Person_B_HeartRate

Heart rate (beats per minute) of Person B

Person_C_HeartRate

Heart rate (beats per minute) of Person C

Person_A_SpO2

Blood oxygen saturation (%) of Person A, indicating oxygen levels in the bloodstream

Person_B_SpO2

Blood oxygen saturation (%) of Person B

Person_C_SpO2

Blood oxygen saturation (%) of Person C

Person_A_Temp

Body temperature (°C) of Person A

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Person_B_Temp

Body temperature (°C) of Person B

Person_C_Temp

Body temperature (°C) of Person C

The dataset contains altitude readings from 7,000 m to 8,300 m, tracking the heart rate, oxygen saturation, and body temperature of three individuals: A, B, and C. No names. The dataset is very confusing, so I decide to keep vital parameters separate for each person. I log in to ChatGPT and write the following prompt:

```
Consider the attached dataset, which measures the vital parameters of three
individuals while they scale Everest.
Create a new worksheet for each person named A, B, and C, each with data
extracted from the provided worksheet.
Format the dataset in the .xlsx format.
```

I also upload Mark's file using the ChatGPT feature to upload files. ChatGPT then provides me with the updated Excel file to download, which I do. As expected, the new file has one worksheet for each person. Figure A-3 shows the generated output for Person A.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure A-3. The Excel file generated by ChatGPT with all the individuals separated

The values for each person are challenging to interpret; thus, I ask ChatGPT to draw a line chart for each metric with the following prompt:

```
In each worksheet, draw a line chart for each metric illustrating how the metric
changes with altitude. Format the resulting workbook in the .xlsx format.
```

I download the file generated by ChatGPT and open it. I note something strange related to Person C. The file contains exactly one chart for each metric and person, nine charts in total. But Person C presents something unusual. I examine Person C's three charts (Figure A-4): the heart rate is increasing, while the temperature and oxygen saturation are decreasing.

Figure A-4. The vital parameters for Person C

I need to look into this further. I absolutely need to know if among these people, the values of some, and perhaps Person C, have dropped far enough below a certain

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

threshold to prevent the continuation of the climb. This would explain why the data stops at 8,200 meters and does not reach the top of Everest, which is officially 8,848 meters high. I ask ChatGPT to classify the vital parameters based on its knowledge to understand whether, at a certain altitude, they are critical or not. I use the following prompt:

```
In each worksheet, add a column that establishes if the vital parameters are
critical (below a certain threshold that you identify as critical) or not.
Format the resulting workbook in the .xlsx format.
```

ChatGPT generates a new file with a new column (Figure A-5).

Figure A-5. The file generated by ChatGPT with a new column called *Status*

The labels given by ChatGPT seem wrong. If a person had critical values starting from an altitude of 7,500 meters, why did they continue climbing? So I ask ChatGPT to provide some clarification about the criteria used to perform the classification. I write the following prompt:

```
Which criterion did you use?
```

ChatGPT provides the following answer:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Here are the criteria used to determine whether the vital signs are "Critical" or "Normal" in the "Status" column for each individual:

Heart Rate

- Normal range: 40 to 120 beats per minute (bpm)
- Critical if:
  - Below 40 bpm (bradycardia)
  - Above 120 bpm (tachycardia)

SpO2 (Blood Oxygen Saturation)

- Normal range: 95–100%
- Mild concern: 90–94%
- Critical if:
  - Below 85%, indicating severe hypoxemia

Body Temperature

- Normal range: 36.1–37.2°C
- Tolerance range set: 35.0–38.0°C (to account for altitude-induced changes)
- Critical if:
  - Below 35.0°C (hypothermia)
  - Above 38.0°C (fever/hyperthermia)

Final Status Logic:

An entry is marked "Critical" if any one of the three metrics is outside its acceptable range. Otherwise, it's marked "Normal".

The problem with this classification is that it does not consider climbing in extreme conditions for all vital signs. So I ask ChatGPT to also consider this aspect through the following prompt:

Consider also a logic for climbers working in extreme conditions.

ChatGPT revises its logic as follows:

Revised Criteria for Extreme Altitude (Above 5500m)

Heart Rate

- Elevated heart rate is expected due to lower oxygen.
- Revised critical range:
  - Below 50 bpm or above 130 bpm

SpO2

- At high altitude, SpO2 normally drops.
- Revised critical threshold by altitude band:
  - 5500–7000m: Critical if SpO2 < 75%
  - 7000–8000m: Critical if SpO2 < 70%
  - Above 8000m (Death Zone): Critical if SpO2 < 65%

Body Temperature

- Cold exposure is extreme, but hypothermia must still be flagged.
- Critical if:
  - < 34.5°C (moderate/severe hypothermia)
  - > 38.0°C (infection/heat stress even in cold)

Altitude-Aware Status Logic

For each row:

1. Determine altitude band.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

2. Apply revised SpO2 threshold.
3. Check if any parameter falls outside adapted critical ranges.
4. Label as "Critical" if any one parameter is outside the acceptable range; otherwise "Normal".

Then it asks me if I want to generate the updated file. I write okay. Then I download the updated file generated by ChatGPT. Now the Status column seems correct (apart from the last line, which is classified as normal). Persons A and B are always classified as normal, while Person C begins to be critical at an altitude of 8,100 m (Figure A-6).

Figure A-6. The revised file generated by ChatGPT with the corrected values for the *Status* column

I close the Excel file and begin reviewing the other files. I find a text file. I open it and read the following text: “I left our plate at 8,212 meters, right there, hoping that one day I will be able to come back and get it.”

Mark did not specify what this “plate” was, nor did he label who each person was. But I start to suspect that one of the people might be Mark. I don’t know for sure, but I can’t ignore it. I close the laptop, and for the first time in a while, I start packing my gear.

I’m going back. To the mountain. To the truth.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Activity 2: Creating Templates

The air smells different at Base Camp at 5,364 meters. The world around is silent on the slopes of Everest.

I carry Mark's old laptop in my backpack, wrapped in waterproof fabric. I've also packed a satellite uplink modem to guarantee intermittent internet access during the climb. I'll use it only in case of emergency and when I need ChatGPT's help with calculations. At Base Camp, I meet the two guides who will take me to the summit: Rajan and Tenzing. We all wear health-monitoring sensors to track vital signs and other physiological data. Here at Base Camp, they're both known to be highly experienced. Rajan doesn't talk much. He watches the mountain like it might speak to him first. I really like him. From the way he looks at the mountain, it seems like he's deeply in love with it. And that makes me feel close to him.

Tenzing, on the other hand, is all smiles and charm. He shakes hands like a man with nothing to hide. But I feel something under the surface between them, something like a tension. For now, I keep it to myself. Instead, I get to work. We also have Sherpas with us who will help us carry our gear to the summit.

I haven't asked them if they know Mark yet. I want to figure something out for myself.

I want to keep track of all vital parameters during the climb to make sure that everything is under control. I open ChatGPT and start a conversation by asking it to generate a template with the vital parameters for Rajan, Tenzing, and me. I write the following prompt:

```
I want to create an Excel template. I'll provide you with all the options to
include in the template, and when I have finished, you'll format it as an
.xlsx file.
```

In the same chat, I write the following instruction for ChatGPT:

```
Create a single worksheet containing 3 worksheets, named Me, Rajan, Tenzing, each
with the following columns:
- Altitude
- HeartRate
- SpO2
- Temperature
Format each table as an Excel table and name with the name of the sheet.
In each sheet add 3 line charts, each representing the Altitude vs the other
column
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

> Add 10 empty rows to the table.

ChatGPT generates the file. After downloading it, I open it in Excel and obtain the result shown in Figure A-7.

Figure A-7. The template generated by ChatGPT designed to keep track of all the vital parameters for Rajan, Tenzing, and me

I still have some time before leaving Base Camp, so I extend my template with information related to the different altitudes.

Climbing Everest is a staged ascent, with four major camps between Base Camp and the summit. To prepare, I decide to prefill my worksheets with the altitudes of each camp, from Base Camp through Camp 4. But I can't recall the exact elevations. So I turn to ChatGPT for help, asking it to extract the correct altitudes and populate the Altitude columns in my tables.

> Fill the column Altitude in each table with five rows, each with the altitude of one of the five camps from the base camp to the Everest peak. Extract the required information from the web. Format the resulting table as an Excel Table. Keep the original table layout and style.

ChatGPT generates a new Excel file. I download it and save it as AppendixA/my_health_log.xlsx. When I open it, I observe that altitudes have been added to each table (Figure A-8).

For now, my work is done. I shut down the computer quickly to preserve the battery, and try to get a little rest. The tent is quiet, but my mind isn't. I keep thinking about what tomorrow might bring and what truths the mountain still hides.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Just as I'm drifting into a shallow calm, I hear footsteps crunching the snow outside.

"May I come in?" It's Rajan's voice.

"Of course," I answer.

Figure A-8. The template generated by ChatGPT populated with the altitudes of the five camps

He steps in and sits down, quiet and composed. His eyes drift to the laptop I've just turned off.

"Why did you bring a computer with you?" he asks.

"To track our vital signs," I say. "Because I'm afraid something might go wrong."

"Why would it?"

"Because it already has."

He studies my face, curious. I see it in his eyes: something flickers. Something almost...familiar.

And so I tell him everything about Mark and what I found in the data.

As I speak, Rajan listens attentively, still. When I finish, he looks at me, and for a moment, I see a softness in his expression, like he knew Mark. Like he was there. But he

says nothing.

So I ask: “Did you know him?”

Rajan doesn’t answer. He just looks away, and the silence between us grows louder than any storm.

Activity 3: Dataset Enrichment

Several weeks have passed since the start of the climb, and we’re finally at Camp 4, the last camp before the final climb. We’re at 8,000 meters. Things are going very well between us, although to be honest, Tenzing is a bit arrogant. He always wants to be right. The other day, while we were crossing a ladder between the glaciers, he almost made me fall off, simply because he rushed me since we were behind schedule. The ladders connect the two overhanging edges of a fracture in the glacier, and if you don’t cross them slowly, you risk falling into the fracture. Rajan, on the other hand, is very quiet and is always very kind to me. I don’t want to exaggerate, but perhaps something more than friendship is developing between Rajan and me.

I’m in front of the computer, and I just recorded the new data in my Excel file titled AppendixA/my_health_log_8000.xlsx. I have in front of me the first worksheet of the file, shown in Figure A-9. I have collected data up to 7,920 meters.

Figure A-9. The Me worksheet in the my_health_log_8000.xlsx dataset

The workbook contains three worksheets, one for Rajan, one for Tenzing, and one for me. For each of us, I’ve populated the data collected at different altitudes relating to heart rate, SpO_2, and temperature. I’ve highlighted the altitudes corresponding to the five camps in orange.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Looking at the data, it seems our vital signs are slowly decreasing, as is normal with increasing altitude. A doubt creeps in. Something doesn't add up, or rather, it's something I've seen before. I try to understand better. I open and look at the AppendixA/mark_health_log.xlsx file (Figure A-10).

There's something similar between this file and my file, but I can't figure it out.

I decide to use ChatGPT to clarify the situation. My goal is to determine if there's a similarity between Rajan's or Tenzing's vital signs and those of Persons A, B, and C in Mark's file. If I find a similarity in the data, maybe Tenzing or Rajan was with Mark.

I activate the satellite connection and connect to chatgpt.com. After logging in to my account, I select o4-mini as the working model from the drop-down menu at the top left of the ChatGPT screen. This model offers advanced reasoning. I click the + symbol and upload the two files, mark_health_log.xlsx and my_health_log_8000.xlsx. The connection is extremely slow, but I wait patiently.

Figure A-10. The Person A worksheet in the mark_health_log.xlsx dataset

Once the files are loaded, I try to merge the two datasets to compare them. I type the following command for ChatGPT:

```
Combine the two attached datasets by appending the tabs of the second dataset to
the first one. Make the output dataset downloadable as an Excel file.
```

I download the file generated by ChatGPT. Figure A-11 shows the results. Not bad, although the data wasn't formatted as tables, and I lost the graphs.

At this point, I want to compare Tenzing's data with that of Persons A, B, and C. To make the comparison, I ask ChatGPT to calculate the Euclidean distance between Tenzing's data and that of each person. Euclidean distance is the shortest path

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

separating two points in space. Essentially, it tells how close or far apart two entities are. The smaller this distance, the more similar the points are. Ideally, two identical points have a Euclidean distance of zero.

I type this prompt for ChatGPT:

```
Add a new tab to the file named Tenzing Analysis where you compute the average
Euclidean distance between Tenzing's data and the other persons' data
(A, B, and C), using only the altitudes that are exactly the same in both
datasets. Specifically: Identify the altitudes that are common to both
datasets. For each common altitude, extract the corresponding [HeartRate,
SpO2, Temperature] values from both datasets. Compute the Euclidean distance
between these vectors at each common altitude using the appropriate Excel
formula. Return the average of these distances for each person (A, B, C) and
Tenzing. Make the result downloadable as an Excel file.
```

I download the new workbook generated by ChatGPT. The workbook contains a new worksheet, called Tenzing Analysis, which contains the results of the comparison between Tenzing and Persons A, B, and C.

Figure A-12 shows the results of the comparison between Tenzing and Person A.

Figure A-11. The new workbook with one worksheet for each person

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure A-12. The comparison between Tenzing and Person A

I'm not entirely satisfied with the formatting, so I ask ChatGPT to fix it, using the following prompt:

```
Reorganize the Tenzing Analysis Formulas sheet so that there is a column for
each person, and at the bottom of each column, compute the average value.
```

I download the new workbook generated by ChatGPT. Figure A-13 shows the result.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure A-13. The rearranged worksheet with the comparison between Tenzing and each person

I notice something truly surprising. The Euclidean distance between Tenzing's data and Person B's data is very small (an average Euclidean distance of about 0.34), so Tenzing could be Person B in Mark's dataset.

I try repeating the operation for Rajan. I type the following prompt for ChatGPT:

```
Repeat for Rajan.
```

I download the new workbook generated by ChatGPT. Figure A-14 shows the result.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure A-14. The rearranged worksheet with the comparison between Rajan and each person

I'm shocked. Rajan's data is very similar to that of Person C in Mark's dataset (an average Euclidean distance of about 0.12), while Tenzing's is similar to Person B. I'm starting to suspect that Rajan and Tenzing were with Mark on that climb and that at some point during the climb, Rajan felt ill.

While I'm thinking about what to do, someone knocks on the tent.

"Come in!" I say.

"Helena!" Rajan greets me, clearly distressed. "Tenzing has disappeared."

Activity 4: Defining Macros

Two hours have passed since Tenzing disappeared, and we still can't find him. I need some rest, so I return to my tent while the rest of the team continues the search. Rajan is coming with me.

I turn on the computer. My goal is to calculate how much longer Tenzing can survive out there. I open the file I've been working on, AppendixA/my_health_log_8000.xlsx, then I start the satellite connection and connect to chatgpt.com. After logging in with my credentials, I type the following prompt for ChatGPT:

```
Consider the attached dataset. Tenzing is missing on Mount Everest and is
probably in trouble. Using the data in the Tenzing worksheet, create a
macro that estimates how long Tenzing can survive at the last available
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
altitude, based on his vital parameters recorded in the dataset and the
knowledge you have about survival in critical environmental conditions.
Make the formula dependent on the last seen values.
```

I click the + button to upload the file my_health_log_8000.xlsx and press Enter. My goal is to use ChatGPT to generate an Excel macro that simulates Tenzing's survival time, based on the latest vital signs I have available.

Before generating the macro, ChatGPT provides some interesting information, including the following:

```
To estimate how long Tenzing can survive at the last known altitude (7900
meters), we'll use a simple model based on these assumptions:
Key Survival Factors at High Altitude:
  1. SpO2 below 80% is critical – risk of unconsciousness or death increases.
  1. Heart Rate indicates exertion – >90 bpm at altitude suggests high stress.
  1. Core Temperature dropping below 35°C (hypothermia threshold) is dangerous.
We'll create a macro that:
  • Takes the last known values.
  • Estimates a survival time (in hours) based on degradation trends in SpO2 and
    temperature.
  • Reduces time if SpO2 or temperature is below thresholds.
```

The proposed model is relatively simplified, but it's fine for my quick analysis. ChatGPT also generates the following macro:

```
Sub EstimateTenzingSurvivalTime()
    ' Define the worksheet to work with
    Dim ws As Worksheet
    Set ws = ThisWorkbook.Sheets("Tenzing")

    ' Find the last row of data in column A (Altitude)
    Dim lastRow As Long
    lastRow = ws.Cells(ws.Rows.Count, "A").End(xlUp).Row

    ' Declare variables to hold the last known values
    Dim altitude As Double
    Dim heartRate As Double
    Dim spO2 As Double
    Dim temperature As Double

    ' Read the last known vital parameters from the worksheet
    altitude = ws.Cells(lastRow, 1).Value    ' Altitude (column A)
    heartRate = ws.Cells(lastRow, 2).Value   ' Heart Rate (column B)
    spO2 = ws.Cells(lastRow, 3).Value        ' SpO2 (oxygen saturation, column C)
    temperature = ws.Cells(lastRow, 4).Value ' Core body temperature (column D)
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
    ' Start with a base survival time assumption (in hours)
    Dim baseHours As Double
    baseHours = 12 ' Base estimate under normal but high-stress conditions

    ' Adjust survival time based on oxygen saturation levels
    If spO2 < 80 Then
        baseHours = baseHours - 6 ' Critically low oxygen, major reduction
    ElseIf spO2 < 85 Then
        baseHours = baseHours - 3 ' Very low oxygen
    ElseIf spO2 < 90 Then
        baseHours = baseHours - 1 ' Mild hypoxia
    End If

    ' Adjust survival time based on temperature (hypothermia risk)
    If temperature < 35 Then
        baseHours = baseHours - 3 ' Hypothermic condition
    ElseIf temperature < 36 Then
        baseHours = baseHours - 1 ' Slightly low body temp
    End If

    ' Adjust survival time based on elevated heart rate (indicating stress)
    If heartRate > 100 Then
        baseHours = baseHours - 2 ' Extreme stress
    ElseIf heartRate > 90 Then
        baseHours = baseHours - 1 ' High but not critical
    End If

    ' Ensure the estimated survival time is not negative
    If baseHours < 0 Then baseHours = 0

    ' Display the final estimated survival time in a message box
    MsgBox "Estimated survival time at " & altitude & " m " & _
    Round(baseHours, 1) & " hours", vbInformation, "Survival Estimate"
End Sub
```

To apply the macro, I return to my my_health_log_8000.xlsx file, select the Developers tab from the Ribbon, click Macro, and enter TenzingSurvival Time as the Macro Name in the dialog box that opens. I click the + symbol. In the editor that opens, I paste the macro generated by ChatGPT. I click Save and then Yes in the dialog box. After my macro is saved, I click the triangular “Play” button (Run Sub/UserForm) in the VBA Editor to run the macro. The dialog box shown in Figure A-15 opens.

I click OK to close the dialog box. My analysis suggests Tenzing can survive for three hours, and he’s already been missing for two.

“We don’t have much time,” I tell Rajan.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

“What does it matter?” Rajan replies. “We’ll continue our climb, even if we don’t find him.”

Rajan’s words chill my heart. I can’t believe he can say those words.

“No!” I say. “We can’t let him die. We must find him at all costs. The summit of Everest isn’t worth a human life.”

Figure A-15. The dialog box showing the result of the macro

Rajan looks at me, petrified. Then he lowers his eyes and, looking at the ground, begins to speak softly.

“It was all three of us: Mark, Tenzing, and I. We were one step away from the summit of Everest. We were the first to take an alternate route, and our morale was sky high. But then I started to feel sick. I couldn’t continue climbing. Tenzing left me there and reached the summit. Then he went back down to the valley, alone, and took all the credit. He made the newspapers and became famous for discovering that alternate route.”

“And Mark?” I ask him, shaken.

“Mark stayed with me and saved my life. He took a plate with our names engraved on it that he had brought to leave at the summit and left it there at the highest point we

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

reached, promising that one day we would return for it to climb to the top."

By now, I have tears in my eyes. I look at Rajan and whisper, "We have to find Tenzing. We have to do it for Mark."

"No," Rajan replies. "It's too late now."

I look at him questioningly.

"Tenzing fell into a ravine. I saw it with my own eyes."

"Did you push him?" The words are out of my mouth before I know they are there.

He's not answering me.

For a moment, I can't move, I can't think, then the shock turns into something hot and violent. The affection that was building inside me for him shatters, replaced by a fury that burns like hot coals.

I grab his arm, my fingers digging in like claws. "Take me there immediately," I say, my voice cold as steel. "Now."

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Epilogue

As soon as we get close to the ravine, we see that our team has found Tenzing and called a rescue helicopter. The rescue team has already pulled Tenzing to safety. He's alive, though not in great shape. The rescue team says he'll probably recover. We decide to descend into the valley and stop before reaching the peak.

Looking around absent-mindedly, I see a folded piece of paper on the ground in the snow. I pick it up and put it in the pocket of my jacket. I'm too confused and angry to read it just then. All my hope of finding Mark is gone. I'm devastated.

The weight of it crushes me. I can barely process it. My disappointment in Rajan cuts deeper than the cold. Once we reach the valley, I'll make sure the authorities hear every detail. And then he'll face exactly what he's earned.

Appendix B. Searching for the Truth

This appendix is the second in a three-part series where I introduce a teaching approach I call *applied fiction*, a new technique to learn through storytelling. In applied fiction, the narrative is a vehicle for teaching real concepts and skills.

The short novel unfolds across three appendices, and each part is paired with hands-on activities tied directly to the book's concepts. You can read all three for the whole arc, or dip into just one and still get a complete set of activities—perfect if you want to focus on skill-building without committing to the entire story.

This appendix reviews concepts from Part II, "Data Science with Copilot".

The Story So Far

Helena, an experienced climber, learns that her brother Mark, once a guide in the Himalayas, has mysteriously disappeared in Nepal after returning home changed and unsettled. Left with his laptop, Helena uncovers cryptic data suggesting that a past Everest expedition went wrong, with one climber showing a critical physical decline near the summit. Determined to uncover the truth, she returns to Everest herself, accompanied by two guides, Rajan and Tenzing. As the ascent progresses, Helena discovers unsettling similarities between her companions and the unidentified climbers in Mark's data, revealing that both men were with Mark during his final climb. The truth slowly emerges: Mark sacrificed his chance at the summit to save Rajan, while Tenzing abandoned them and claimed glory alone. When Tenzing later disappears during the climb, Rajan confesses his role in the past and hints at darker actions. Although Tenzing is eventually rescued alive, Helena abandons the summit and resolves to expose the truth, disillusioned by betrayal and haunted by Mark's unresolved fate.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Prologue

I'm still here, in Nepal, in a hotel room in Kathmandu with that silent anger weighing on me. I think about how I could have fallen for Rajan and how angry I feel toward him. I just want to go home and forget everything, but I can't. My heart is full of emotions.

While I'm still immersed in these thoughts, my gaze falls on Mark's computer. The real reason I came here comes back to me: to find my brother. And so I decide to put my angry feelings aside. I gather my courage and turn on Mark's computer again.

I look at the various files I've already seen several times. I start opening them, and with a deep breath, I begin examining them one by one. To be honest, I doubt I'll find anything interesting, but I'll give it a try anyway. Maybe I'll find something useful, something that will help me find him.

Activity 1: Preparing the Dataset

I sit down and get started. Browsing through the various folders, I find a series of scattered files. One in particular stands out, called *AppendixB/transactions.xlsx*.

I take a deep breath, and then I begin. I open the file *transactions.xlsx* and see a screen similar to the one shown in Figure B-1.

Figure B-1. A screenshot of the first 20 lines of transactions.xlsx

Another file in the same folder describes the table fields:

TransactionID

A string identifier for a transaction

Date

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

The transaction date

TouristName

The full name of the client participating in the Everest expedition

AmountUSD

The monetary amount charged for the transaction, expressed in US dollars

Guide

The guide assigned to the transaction

TourType

A categorical description of the expedition phase or service provided

Notes

Free-text operational annotations related to the transaction, such as logistics, weather conditions, permits, or medical checks

The file is very messy; I see several duplicates and formatting issues. I decide to ask Copilot for help sorting out this data. First, I want to remove the duplicates. I access the Copilot chat by clicking the corresponding button in the Ribbon's Home tab. Next, I select "Agent mode" from the Tools button in the Copilot chat and type this prompt:

```
Consider the Transactions table. Highlight duplicates in the column TransactionID
```

Copilot highlights all repetitions. There are several, as shown in Figure B-2.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure B-2. Some duplicates highlighted by Copilot (rows 17 and 18)

I read Copilot's explanation to verify the output's correctness. In particular, I focus on the "Show details" section, which explains how Copilot reasoned:

```
What I did
  • Located the Transactions table and identified the TransactionID column.
  • Applied a Custom conditional formatting rule to A2:A101:
    • Formula: COUNTIF($A:$A, A2)>1
    • Format: yellow fill, bold, black font.
  • Auto-fit the entire column A to ensure highlighted cells are fully visible.
Verification
  • Read back the TransactionID values and computed duplicates to confirm the
    rule targets the right set:
    • Found 10 duplicated IDs: E00018, E00052, E00059, E00065, E00073, E00075,
      E00077, E00079, E00086, E00087.
  • Confirmed the applied range: Transactions!A2:A101.
Notes
  • The formula uses an absolute reference to column A and a relative reference
    to the first data cell (A2), so it evaluates correctly across the entire
    column range.
  • The rule will continue to work as long as new data is added within the
    applied range; if you extend the table, reapply the rule to cover the
    new rows.
```

Copilot's reasoning is correct, thus I can proceed. Although the Amount USD column contains different values for a duplicate entry, I decide to remove duplicates anyway. So, I ask Copilot to remove the duplicates. I use the following prompt:

```
Consider the Transactions table. Remove duplicate entries based on TransactionID
```

As a result, Copilot removes all the duplicates, as shown in Figure B-3.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure B-3. The first 20 rows of the dataset without duplicates (TransactionID E00018 appears just once)

I evaluate the results by reading the Copilot explanation. In particular, I focus on the following piece of output generated by Copilot:

```
What I did (quickly):
  • Checked the Transactions table's TransactionID column, identified duplicates,
    and deleted the later occurrences (keeping the first).
  • Verified that all remaining IDs are unique and tidied the table columns for
    visibility.
```

This output explains that Copilot has removed the second duplicate.

I'm starting to see something strange, but I still don't understand it. The file still contains too much noise. I decide to standardize the data into a single format.

First, I'll align the dates. I write the following prompt for Copilot:

```
Consider the Transactions table. Convert all dates in the Date column to format
YYYY-MM-DD
```

I get the result shown in Figure B-4. I have my local language set to Italian, so Copilot has interpreted the dates using the Italian format.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure B-4. The first 20 rows of the dataset with the Date column properly formatted

As usual, I review Copilot's description of output—specifically, I look at the "What I did" section:

> What I did:
> - Standardized the Date column to ISO format. Where locale differences made parsing inconsistent, I reconstructed ISO dates from the table's provided dataset and applied them uniformly.
> - Confirmed the result by reading back the displayed text of every cell and checking that it matches YYYY-MM-DD.

At this point, I decide to replace the guides' names with the names I discovered in my previous analysis (see Appendix A for more details). Person B is Tenzing, Person C is Rajan, and Person A is Mark. I write this prompt for Copilot:

> Consider the Transactions table. In the column Guide, replace Person B with Tenzing, Person C with Rajan, and Person A with Mark.

I get the result shown in Figure B-5.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure B-5. The first 20 rows of the dataset with guide names changed in the Guide column

When I finish, I feel a tightening in my stomach: when Tenzing is the guide, the payments from climbers are unusually high. For example (see Figure B-5 for comparison), in row 6, I notice a payment of $5,204.57 from Sam Kim, and the guide was Tenzing. Two days before (row 5), another high payment of $5,533.91 and the guide was again Tenzing. In row 8, by contrast, there is a payment to Mark of $3,660.85, which is much less than the amounts paid to Tenzing. Similarly, Rajan received about $3,000 in each of his transactions (e.g., rows 17 and 18).

I stop. My hands are shaking a little. Something doesn't add up here. There's only one way to confirm what I think I'm seeing: use statistics.

Activity 2: Basic Statistics

After discovering those suspicious payments, I feel even more determined to figure out what Tenzing was up to. So I move to the next step: deeper analysis. I see that certain tourists were paying exorbitant amounts, and that these payments were always associated with Tenzing. I'm starting to think that Rajan and Mark might have been unwittingly involved in something shady—or that they had discovered something. Maybe things on Everest didn't happen exactly the way Rajan told me they had. Maybe Rajan and Tenzing had an argument on Everest, and then something happened. I get even angrier, but I transform that anger into focus. Every number I put together tells me I'm getting closer to the truth.

I take a sip of tea, and then I begin. My goal is to compare the payments made to Mark, Rajan, and Tenzing to determine whether there is a hidden money flow behind the expeditions to Everest.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

First, I calculate the central tendency measures. Using the Agent mode, I write this prompt for Copilot:

```
Consider the Transactions table. Consider the columns Amount USD and Guide.
Calculate the mean, median, and mode using Excel formulas when the Guide
is Tenzing. Repeat when the Guide is Rajan. Repeat when the Guide is Mark.
Add the results as a table to a new sheet
```

Copilot then adds a new sheet containing the calculated metrics, as shown in Figure B-6.

Figure B-6. Basic statistics for Tenzing, Rajan, and Mark

Regarding the mode, Copilot adds the following comment:

```
Assumption: MODE.SNGL returns #N/A when no repeated amounts exist—which matches
this dataset.
```

As expected, the average amount for Tenzing is much higher than that for Rajan and Mark. The same applies to the median value. I'm beginning to fear that every climb with him came at a higher price…and it wasn't just money.

To understand how much these values vary, I calculate the dispersion measures. I select cell A1 in the worksheet with the basic statistics. I click the Table Design tab in the Ribbon to retrieve the table name (it's GuideStats in my case). I write the following prompt for Copilot:

```
Consider the GuideStats table. Add the standard deviation column. Extract
information from the Transactions table.
```

Copilot adds a new column in the table, as shown in Figure B-7.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure B-7. The table with the StdDEV column

The section “How it was done” of Copilot’s answer provides more details about the formula used:

```
How it was done:
Inserted a new table column and filled it with:
=STDEV.S(FILTER(Transactions[Amount USD], Transactions[Guide]=[@Guide]))
```

Copilot also adds this detail:

```
Assumption: You wanted the sample standard deviation (STDEV.S). If you prefer
population standard deviation, I can switch to STDEV.P.
```

Tenzing’s standard deviation is very high, meaning that payments vary suspiciously and fluctuate. Rajan and Mark’s standard deviations, on the other hand, appear consistent and transparent.

And suddenly, I feel a knot in my stomach. Tenzing and Rajan’s argument probably wasn’t just on the mountain. Tenzing had something to hide. And, even worse, perhaps Mark’s disappearance is linked to this situation.

Activity 3: Data Exploration and Visualization

At this point, all that’s left to do is build a graph that shows the differences. My goal is to create clear and simple visualizations that show:

1. Total money paid to Rajan vs. Tenzing vs. Mark
2. Payments over time

I decide to create a bar chart showing the pay gaps between Rajan, Mark, and Tenzing. I give Copilot the following prompt:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
Consider the GuideStats table. Insert a column chart comparing the average
Amount USD for each Guide. Add chart title "Average Payments by Guide".
Label the y-axis "Amount (USD)". Remove gridlines.
```

Copilot generates the graph shown in Figure B-8.

Figure B-8. The column chart generated by Copilot

From the chart, it's very clear that Tenzing's average amount is much higher than Rajan's and Mark's. If necessary, I will use this chart in court to demonstrate Tenzing's shady dealings.

To show Tenzing's payment trend, I build a line chart. I type the following prompt for Copilot:

```
Consider the Transactions table. Insert a new sheet named Payments for 2025 -
Tenzing containing a pivot line chart showing Amount USD over Date when
Guide = "Tenzing". Add chart title "Payment Trends". Label x-axis as "Date"
and y-axis as "Amount (USD)". Remove markers and simplify the layout.
```

Copilot adds a new sheet, called Tenzing Trend, containing a PivotTable with only Tenzing's data and the associated chart (Figure B-9).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure B-9. The worksheet showing Tenzing's Payments for 2025

I repeat for Mark, using the following prompt:

```
Repeat for Mark
```

Copilot creates a new sheet with Mark's data and the chart shown in Figure B-10.

Figure B-10. Mark's Payments for 2025

The trend line contains two strange peaks. I look at the Copilot answer to find a possible explanation. I find this:

```
Assumption: When multiple Mark payments occur on the same date, the chart shows
the daily total (via SUMIFS).
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

This means there were some days when Mark received more than one payment. For Tenzing, this doesn't seem to happen.

I repeat the same instruction for Rajan:

```
Repeat for Rajan
```

Figure B-11 shows the resulting sheet.

Figure B-11. Rajan's Payments for 2025

For Rajan, there are also some days with a peak, meaning that he also received multiple payments in a single day. To be sure of this, I start a new chat in Copilot and select the Analyst modality from the Tools button of the chat. I write the following prompt:

```
Consider the Transactions table. Select only data with Guide = Tenzing. Are there
more than one transaction for the same date?
```

Copilot provides me with the following answer:

```
For the guide Tenzing, there are no dates with more than one transaction in the
Transactions table. Each transaction date for Tenzing is unique.
```

I continue the conversation with Copilot by writing the following prompt:

```
Repeat for Mark
```

I get the following answer:

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
Yes, for the guide Mark, there are dates with more than one transaction in the
Transactions table:
  • Date 45891 has 2 transactions.
  • Date 45970 has 2 transactions.
```

I note that dates are represented as serial numbers. I repeat for Rajan, using the following prompt:

```
Repeat for Rajan
```

I get the following answer:

```
Yes, for the guide Rajan, there is a date (45725) with more than one transaction
in the Transactions table (in fact, three transactions on that date).
```

It seems Mark received more climbing requests than Tenzing, especially in the last period. To confirm this, I write the following prompt for Copilot (in the same chat, using the Analyst modality):

```
Consider the Transactions table. Consider Guide = Tenzing. Calculate the number
of transactions from November to December 2025.
```

Copilot provides the following output:

```
Number of transactions for Tenzing in the last two months: 4
```

I repeat for Mark, using the following prompt:

```
Repeat when the Guide is Mark
```

Copilot provides the following output:

```
Number of transactions for Mark from November to December 2025: 10
```

Just out of curiosity, I ask Copilot to repeat for Rajan:

```
Repeat when the Guide is Rajan
```

Copilot provides the following output:

```
Number of transactions for Rajan from November to December 2025: 4
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Tenzing earns much more in a single transaction, but tourists prefer Mark. I want to check if this is reflected in the monthly earnings. Tenzing earns more money than Mark in a single transaction, but I suspect he earns less monthly because Mark receives more payments. To validate my hypothesis, I ask Copilot to extract the month from the Date column and then build the charts. I start a new chat and write the following prompt, using the Agent mode:

```
Consider the Transactions table. Extract the month from the Date column and add
as a new column. Next, insert a new sheet called Monthly Payments - Tenzing
containing a pivot column chart showing Amount USD over Month when
Guide = "Tenzing". Add a chart title “Monthly Payments". Label x-axis as
"Month" and y-axis as "Amount (USD)". Simplify the layout.
```

Figure B-12 shows the resulting sheet.

Figure B-12. Tenzing’s monthly payments

Over the last three months, Tenzing’s payments progressively decreased. I want to compare his data with Mark’s, so I write the following prompt:

```
Repeat when the guide is Mark
```

Figure B-13 shows the results.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure B-13. Mark's monthly payments

I repeat also for Rajan, using the following prompt:

```
Repeat when the guide is Rajan
```

Figure B-14 shows the results.

Figure B-14. Rajan's monthly payments

Although Mark earned less than Tenzing in a single transaction, each month he earned more because he received multiple payments. Perhaps Tenzing wanted to get rid of Mark to take his share of the tourists as well. I decide to reorganize all these results into a report I will share with a certain person I know.

Activity 4: Data Reporting

To organize my workbook as a report, I first create a new worksheet, which will contain all the main information about my report. I click the + symbol at the bottom right

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

of the window, at the end of all the available worksheets. Next I right-click the new worksheet and select Rename from the drop-down menu. I name the new worksheet Summary. I drag the Summary sheet to the left until it becomes the first worksheet.

To populate the Summary worksheet with a title, a short description, and a table of contents, I start a new chat in Copilot. I don't select any specific modality. I write the following prompt to define the title:

> Propose five titles for a report on transactions involving three guides to the Himalayas. The title should highlight, using a professional tone, that there are dark affairs for the guide named Tenzing.

Copilot proposes the following titles:

> Here are five title suggestions:
>
> 1. "Himalayan Transactions Unveiled: Analyzing Patterns and Unusual Dealings Among Guides, with a Focus on Tenzing"
>   - Highlights the comprehensive analysis and signals special attention to questionable activities.
> 2. "Guiding the Heights: Financial Trends and Shadowed Dealings in the Himalayas—A Closer Look at Tenzing"
>   - Suggests both legitimate and suspicious trends, with Tenzing as a focal point.
> 3. "Beyond the Summit: Transactional Insights and Alleged Irregularities in Himalayan Tours Led by Tenzing"
>   - Implies a deeper investigation into irregularities.
> 4. "Three Guides, Many Journeys: Financial Transparency and the Enigma of Tenzing's Transactions"
>   - Balances transparency with the mystery surrounding Tenzing.
> 5. "From Peaks to Shadows: A Forensic Review of Himalayan Tour Transactions and the Dark Affairs of Tenzing"
>   - Uses strong language to emphasize the seriousness of the alleged "dark affairs."

I like title number 3, but I decide to remove the first part. Thus, the title of my report becomes: "Transactional Insights and Alleged Irregularities in Himalayan Tours Led by Tenzing."

Now, I want to add a short description for my report. I write the following prompt for Copilot:

> Based on the columns in the Transactions table, provide a short description of the dataset

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Copilot writes the following output:

> Dataset Description: Himalayan Tour Transactions
> This dataset records individual financial transactions for guided tours in the Himalayas, involving three guides. Each entry includes:
> - TransactionID: A unique identifier for each transaction.
> - Date: The transaction date (in serial number format).
> - TouristName: The name of the tourist participating in the tour.
> - AmountUSD: The amount paid in US dollars for the tour or activity.
> - Guide: The assigned guide for the tour (one of three guides).
> - TourType: The specific type of tour or activity undertaken.
> - Notes: Additional remarks or status updates related to the transaction.

Finally, I want to add a table of contents. I write the following prompt:

> List the worksheets in this workbook

Copilot writes the following output:

> 1. Summary
> 2. Transactions
> 3. Guide Stats
> 4. Payments for 2025 - Tenzing
> 5. Payments for 2025 - Mark
> 6. Payments for 2025 - Rajan
> 7. Monthly Payments - Tenzing
> 8. Monthly Payments - Mark
> 9. Mouthly Payments - Rajan

I copy all the information to the Summary worksheet. I also adjust the font and add the link to each worksheet in the table of contents. Figure B-15 shows the results.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure B-15. The Summary worksheet

The report is ready. There is just one person who needs to see it. I absolutely must speak to Rajan.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Epilogue

I put aside my anger and decide to visit Rajan, who has been put in prison after the incident with Tenzing. I take Mark's computer with me. As I enter the interview room, I look him in the eye and tell him everything I've found. I show him the files, explain the suspicious payments, and show him the charts I've created. Rajan listens to me and finally nods, with a mixture of relief and sadness. He confesses that yes, he had discovered Tenzing's business dealings, and that their argument on Everest had been truly heated. But he hadn't pushed Tenzing into the ravine. Tenzing had fallen during that argument, and it was clear that Rajan had somehow been framed or misunderstood. Eventually, when everything comes to light, the truth begins to emerge.

"Why didn't you tell the truth right away?" I ask him.

"Because I didn't have proof," he replies, his eyes downcast. "I knew something was wrong, but I wasn't sure."

"How long has this been going on?"

"For a long time, at least two years."

"Did you know that Mark suspected something?"

"No. We didn't speak about this."

"Now, with this evidence, we will be able to prove the truth before the judge, and, finally, you can go free."

I see a tear roll down Rajan's face. I start crying too. I look into his eyes. He looks back at me. The glass separates us, and I can't hug him. All my anger toward him has finally disappeared. I feel great relief.

I place my hand on the glass, and he, in response, places his on the other side. It's as if we've embraced. I finally feel a strong emotion between us that goes beyond friendship.

Appendix C. The Showdown

This appendix is the third in a three-part series in which I introduce a teaching approach I call *applied fiction*. Applied fiction uses narrative to teach real concepts and skills. As you follow the story, you'll naturally absorb the how-to knowledge woven into it, learning by experiencing rather than by rote instruction.

The short novel unfolds across three appendices, with each part paired with hands-on activities tied directly to the book's concepts. You can read all three for the whole arc, or dip into just one and still get a complete set of activities—perfect if you want to focus on skill-building without committing to the entire story.

This appendix reviews concepts from Part III, "Other AI Tools for Excel".

The Story So Far

In Episode 1, Helena, an experienced climber, discovers mysterious health data on her brother Mark's laptop, linked to a failed Everest expedition that stopped short of the summit. By analyzing the files with the help of ChatGPT, she realizes that one of the three climbers showed dangerously critical vital signs at extreme altitude. Suspecting that this data hides the truth about Mark's disappearance, Helena returns to Everest with two guides, Rajan and Tenzing. As the climb progresses, further analysis reveals that both men were part of Mark's final expedition. The truth slowly emerges: Mark gave up his chance to reach the summit in order to save Rajan, while Tenzing abandoned them and later took full credit for the ascent. When Tenzing disappears during the current climb, suspicion falls on Rajan, who appears to have pushed him into a ravine. Tenzing, however, is ultimately rescued. Shaken by the events, Helena abandons her attempt on the summit, choosing to place human life above ambition.

In Episode 2, still in Nepal, Helena continues her investigation by examining Mark's financial files. Using Copilot to clean, analyze, and visualize transaction data, she uncovers suspicious payment patterns linked to Tenzing, whose earnings are unusually high and inconsistent compared to those of Mark and Rajan. Through statistical analysis and charts, Helena finds evidence of questionable business dealings that may explain the conflicts between the climbers. Determined to uncover the full truth, she compiles the findings into a detailed report and confronts Rajan, who is now in prison. Rajan admits that he had discovered Tenzing's illicit activities and that they had argued

violently on Everest, but he firmly denies pushing Tenzing into the ravine. With concrete evidence finally available, Helena vows to bring the truth to light before the authorities, in the hope of clearing Rajan's name. As justice begins to take shape, Helena and Rajan share a quiet emotional reconciliation, bound by loss, truth, and the hope of a future beyond betrayal.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Prologue

Today is a great day. Rajan has finally been released from prison, thanks to the evidence I presented in court. I'm waiting for him outside the Kathmandu prison. It's very cold, so I'm wearing the same jacket I wore during our ill-fated Everest ascent. Finally, after so much suffering, I'll be able to hug him. When he comes out of the prison door, he looks tired but happy. I run to hug him, and he hugs me back. There's no longer that cold glass separating us. It's as if this embrace will last a lifetime. He smiles at me and holds me tightly. I feel indescribable joy.

"Thank you," he whispers softly. "Without you, I never would have gotten out of here."

I smile at him and hug him tighter.

"We have to find Mark," he continues. "I feel like he's still alive."

"Yes," I agree. "But where can we look for him? He could be anywhere, and we have no clue."

For a moment, neither of us speaks. The weight of that uncertainty hangs between us. Then Rajan pulls back slightly, his expression changing, more focused, more alert.

"There is one possibility we haven't faced yet," he says quietly.

I look at him, waiting.

"Tenzing might have hidden Mark somewhere. Mark had more clients, more connections, and Tenzing had too many reasons to make Mark disappear without raising suspicion."

Rajan suddenly stops.

"During my argument with Tenzing," he says, "something fell out of his pocket."

I freeze.

"I think I picked it up," I reply, reaching into my jacket. "I remember putting it in my pocket without thinking."

I pull out a folded piece of paper and open it:

Kh. East
stone house, off trail
yak
~3 km BC
family: N Sh.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

We stare at it for a moment.

“It looks like directions,” Rajan says slowly. “Possibly the location of a Sherpa family’s house. Many live near Base Camp.”

“So if Tenzing hid Mark,” I say, “this could be where he took him.”

Rajan nods.

“Maybe this is the place,” he says. “Come with me. I think I know where to look.”

Activity 1: Data Cleaning with Numerous.ai

A wise old man from the area has given us a list of Sherpa families living near Base Camp. We sit at the small desk in my hotel room in Kathmandu. We quickly write down all the information in the Excel spreadsheet *AppendixC/sherpa_families.xlsx* shown in Figure C-1.

Figure C-1. The data collected regarding the Sherpa families

The dataset contains the following fields:

FamilyID

A unique identifier assigned to each family record.

Family Name

The commonly used name or surname associated with the family.

Village

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

The village or settlement where the family primarily resides. Spelling and level of geographic detail may vary.

Region

The broader geographic region in which the village is located (for example, Khumbu East).

AddressDescription

A free-text description of the household's location. May reference landmarks, trails, or nearby features and is intended primarily for human interpretation.

AltitudeMeters

The altitude in meters.

DistanceFromBaseCampKm

The reported distance, in kilometers, between the household and Base Camp. Values may include approximations or textual annotations (e.g., "~3 km").

HousePosition

The elevation of the household's location measured in meters above sea level.

NearMainTrail

A raw indicator describing whether the house is near the main trekking trail.

AnimalsOwned

A list of animals owned by the family (e.g., yak, cow, goat). Multiple animals may appear in a single cell.

SpeaksEnglish

Indicates whether at least one member of the household is able to communicate in English.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

HouseholdSize

The total number of individuals living in the household.

PrimaryActivity

The main economic or subsistence activity of the household, such as herding, farming, trekking support, or trade.

Having compiled the data quickly, we made a few mistakes, so the first step is to clean the data for subsequent analyses. Rather than overwrite the original values, we choose a safer approach: for each problematic column, we will create a clean counterpart and verify the results before proceeding.

I take control of the computer since I have a lot of recent experience using Excel with AI tools. I decide to use Numerous.ai to clean the data. I perform the following steps:

1. I click cell N1 and write this text: DistanceFromBaseCampKm_Clean. A new table column automatically appears.
2. I select cell N2 and write this text:

```
=NUMBERVALUE(NUMAI(
"Convert to decimal number",
[@DistanceFromBaseCampKm]
))
```

   By default the function NUMAI() returns text, so I have to convert it to a number, using the function NUMBERVALUE().

“Why don’t you use NUMBERVALUE() directly without NUMAI()?” Rajan asks me.

“Because NUMBERVALUE() doesn’t recognize numbers with a decimal point as decimals, so it would convert the number 3.30 to 330.”

“Got it. Thanks.”

I press the Enter key, and after a while, I see the result shown in Figure C-2.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure C-2. The result after running the NUMAI() *function to clean* DistanceFromBaseCampKm

At this point, I add a column called NearMainTrail_Clean starting from cell O1. In cell O2, I type the following prompt to obtain the clean version of the NearMainTrail column:

```
=NUMAI(
"Standardize this value to YES or NO.
Possible inputs include YES, Yes, Y, NO, No, N, Yes with spaces, No with spaces.
Return only YES or NO.",
[@NearMainTrail]
)
```

I get the result shown in Figure C-3.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure C-3. The result after running the NUMAI() *function to clean* NearMainTrail

Next, we need to fix the column SpeaksEnglish. In column P, I add a new column named SpeaksEnglish_Clean and write the following formula in cell P2:

```
=NUMAI(
"Standardize this value to YES or NO.
Possible inputs include YES, Yes, Y, NO, No, N, Yes with spaces, No with spaces.
Return only YES or NO.",
[@SpeaksEnglish]
)
```

I get the result shown in Figure C-4.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure C-4. The result after running the NUMAI() *function to clean* SpeaksEnglish

Finally, I turn to AnimalsOwned, the messiest column of all. In column Q, I add AnimalsOwned_Clean and enter the following formula in cell Q2:

```
=NUMAI(
"Rewrite this list of animals using comma-separated values.
Replace semicolons with commas.
Make all animal names lowercase.
Keep 'none' if no animals are listed.
Return only the cleaned list.",
[@AnimalsOwned]
)
```

I get the results shown in Figure C-5.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure C-5. The result after running the `NUMAI()` *function to clean* `AnimalsOwned`

At this point, the table is clean and ready for analysis. Our task is to determine whether there's anything similar to what's on the note.

"I hope we can extract something meaningful from this," I say to Rajan, lifting a cup of tea and letting the steam warm my face.

"I hope so too," he replies.

He holds my gaze a moment longer than necessary, then lowers his voice, almost as if the data itself were listening.

"I never imagined I would meet the sister of the man who saved my life."

"Neither did I."

For a moment, neither of us speaks. The room is quiet, filled only with the hum of the computer and the weight of what we're about to uncover. Whatever the data reveals, we both know it will change more than just our analysis.

Activity 2: Data Extraction Using PY()

To decrypt the note, we need to extract structured information from the `AddressDescription` column. We decide to use `=PY()`. To generate the Python code, let's ask Copilot for help. I start a new conversation in Copilot and type the following prompt:

```
Use the PY() function provided by Excel. Consider the cells A1:Q21 in the Raw
worksheet, write the code in Python that:
  1. Selects all the table cells in column A1:Q21 using xl() with headers=True
     and stores them in a variable called df
  2. extracts structured information from the column AddressDescription,
     including the house material, where it is (e.g., near to), and other details
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

3. stores all the columns of the original table and the result into a new table

Copilot gives me the following output:

```
import pandas as pd
import re

# Step 1: Load the data from Excel
df = xl("Raw!A1:Q21", headers=True)

# Step 2: Define a function to extract structured info from AddressDescription
def extract_address_info(desc):
    # Example patterns (customize as needed for your data)
    material = None
    location = None
    other = None

    # Try to extract house material
    # (e.g., stone, wooden, terraced, small, old, low, remote, isolated)
    material_match = re.search
        (r'\b(stone|wooden|terraced|small|old|low|remote|isolated)\b',
        desc, re.IGNORECASE)
    if material_match:
        material = material_match.group(0).capitalize()

    # Try to extract location
    # (e.g., near, above, below, north of, behind, off, close to, on, at, by)
    location_match = re.search
        (r'(near|above|below|north of|behind|off|close to|on|at|by)[^,;]*',
        desc, re.IGNORECASE)
    if location_match:
        location = location_match.group(0).strip().capitalize()

    # Other details: anything not captured above
    other = desc
    if material:
        other = other.replace(material, '')
    if location:
        other = other.replace(location, '')
    other = other.strip(',; .')

    return pd.Series({'HouseMaterial': material, 'HouseLocation': location,
        'OtherAddressDetails': other})

# Step 3: Apply the extraction function to the AddressDescription column
address_info = df['AddressDescription'].apply(extract_address_info)

# Step 4: Concatenate the new columns to the original dataframe
result = pd.concat([df, address_info], axis=1)
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
# Step 5: Output the new table
result
```

I check the output, and it looks fine. I follow these steps to run the Python code:

1. I create a new worksheet by pressing the + button at the bottom, next to my Raw worksheet.
2. I start writing =PY(in cell A1. I don't insert the closing bracket. The Python environment has now started.
3. I paste the Python code into the cell.
4. I press Command + Return on my Mac. (On Windows, press Ctrl + Enter.)

I get a single cell with the DataFrame, as shown in Figure C-6.

Figure C-6. The output generated by running =PY()

To see the generated values, I go to the Formulas tab of the Ribbon and select Editor in the Python box. An area opens on the right. I click the Python Output button shown in Figure C-7 and select Excel Value from the drop-down menu that opens.

Figure C-7. The button to click in the Python Editor to select Excel values

At this point, I click the Save button in the Python Editor, and obtain the result shown in part of Figure C-8 (for space reasons, it only shows columns H onwards).

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Figure C-8. The range generated using the function =PY( ) and showing the values directly

"Why don't you convert the data into a table?" Rajan asks me.

"The data shown can't be converted to a table because it's dynamic. I'll just copy the values I need into the main table."

I select the range R1:T21 and copy it by pressing Command + C (if I had Windows, I would have clicked Ctrl + C). I move to the Raw sheet and select cell R1. From the Edit menu, I select Paste Special → Values → OK. I get the result shown in Figure C-9 (the figure shows only the added columns).

Figure C-9. The HouseMaterial, HouseLocation, *and* OtherAddressDetails *fields have been added to the table*

"I think we're finally in a position to predict where Mark is most likely to be," I say, a quiet sense of satisfaction slipping into my voice.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Rajan studies the screen for a moment, then smiles softly. “Let’s hope we’re right,” he says. “And that we get to him in time.”

Activity 3: Prediction with Python and AI

Our goal now is to write a predictor that predicts which row in our table matches the contents of our note. I start a new ChatGPT chat and type the following prompt:

```
Consider the attached dataset. Write a Python function for macOS using xlwings
and OpenAI that predicts if each row meets the following requirements:
Kh. East
stone house, off trail
yak
~3 km BC
family: N Sh..
The prediction output should be YES or NO. The function must read the values from
the worksheet using xlwings, apply the model to each row, and write the resulting
category back into Excel. Call the main function main(). When you use OpenAI
import the OPENAI_API_KEY using this code:
from pathlib import Path def load_api_key(path="openai_key.txt"):
    base = Path(__file__).parent
    abs_path = (base / path).expanduser().resolve()
    with abs_path.open("r", encoding="utf-8") as f:
        return f.read().strip()
os.environ["OPENAI_API_KEY"] = load_api_key()
client = OpenAI()
```

I attach the Excel file to the prompt and run it.

“Why don’t you use Copilot?” Rajan asks.

“Because I want to run the code outside of Excel, and ChatGPT sees the external environment, whereas Copilot is better for the internal environment.”

ChatGPT generates a fairly complex code (I won’t report it here because it gives me an error). I paste it into a new file and save it as *extract.py*. I’ve already stored my OpenAI API key in the *openai_key.txt* file by simply pasting it into the file as is. When I try to run it, however, it gives me this error:

```
Traceback (most recent call last):
  File "extract.py", line 145, in <module>
    main()
  File "extract.py", line 97, in main
    wb = xw.Book.caller() if xw.apps and xw.apps.active else xw.Book()
  File "/Library/Frameworks/Python.framework/Versions/3.8/lib/python3.8
  /site-packages/xlwings/main.py", line 1034, in caller
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
    elif xlwings._xlwindows.BOOK_CALLER:
AttributeError: module 'xlwings' has no attribute '_xlwindows'
```

I copy the error and paste it into the same ChatGPT chat that generated the code. ChatGPT generates some very complex code, which I honestly don't understand. A little annoyed, I type this prompt:

```
Rewrite by keeping the code as simple as possible
```

At this point, ChatGPT generates the following code:

```
import os
import xlwings as xw
from openai import OpenAI
from pathlib import Path

# --- OpenAI key loading (as requested) ---
def load_api_key(path="openai_key.txt"):
    base = Path(__file__).parent
    abs_path = (base / path).expanduser().resolve()
    with abs_path.open("r", encoding="utf-8") as f:
        return f.read().strip()

os.environ["OPENAI_API_KEY"] = load_api_key()
client = OpenAI()

MODEL = os.getenv("OPENAI_MODEL", "gpt-4o-mini")

def predict_yes_no(row_dict):
    prompt = f"""
Decide if the row meets ALL requirements:
- Kh. East
- stone house, off trail
- yak
- ~3 km BC
- family: N Sh..

Row (JSON): {row_dict}

Reply with ONLY: YES or NO
""".strip()

    r = client.chat.completions.create(
        model=MODEL,
        temperature=0,
        messages=[{"role": "user", "content": prompt}],
    )
    out = (r.choices[0].message.content or "").strip().upper()
    return "YES" if out.startswith("YES") else "NO"
```

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

```
def main():
    # Works both from Excel and from terminal on macOS
    try:
        wb = xw.Book.caller()
    except Exception:
        wb = xw.Book("sherpa_families3.xlsx")  # change if your file name differs

    sht = wb.sheets["Raw"] if "Raw" in [s.name for s in wb.sheets] \
    else wb.sheets.active

    data = sht.range("A1").expand("table").value
    headers = [str(h).strip() if h is not None else "" for h in data[0]]
    rows = data[1:]

    # Write predictions into a new column at the end
    out_col = len(headers) + 1
    sht.range((1, out_col)).value = "Prediction"

pres = []
    for row in rows:
        row_dict = {headers[i] or f"col_{i+1}": row[i] for i in \
        range(len(headers))}
        preds.append([predict_yes_no(row_dict)])

sht.range((2, out_col)).resize(len(preds), 1).value = preds


if __name__ == "__main__":
    main()
```

I run it from the terminal with the following command (after using the cd command to move to the folder containing the file):

```
python3 extract.py
```

Once the script has finished running, I reopen the Excel file *sherpa_families.xlsx* and find that a new column has been added to my table, as shown in Figure C-10.

The data leaves no room for doubt. There is only one row marked with the value YES, the third row excluding the header, and it corresponds to Nomi Sherpa. Mark is likely to be at Nomi Sherpa's house. So we make a decision without saying much. We pack up and go to demand answers.

During the journey, we talk quietly, almost to calm our nerves. Rajan is gentle and attentive. Sitting beside him, I suddenly feel that I may have finally found the right person for me.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

When we reach our destination, a small Sherpa village unfolds before us. Simple houses, prayer flags fluttering in the wind, and a silence broken only by distant voices. We stop a man in the street and ask him about Nomi Sherpa.

“Nomi Sherpa and his family live in the house at the end,” he tells us, pointing down the path.

Figure C-10. The new column added by the Python script

We head straight there. A llama stands in front of the door, as if it is expecting us. It feels like a silent confirmation that our suspicions are right. We knock. An elderly woman opens the door and looks at us with hard eyes.

“What do you want?” she asks in Nepali, her tone sharp and unfriendly.

“We’re looking for this man,” I say, showing her a photo of my brother.

Her body stiffens instantly.

“I don’t know this man. I’m sorry.”

She slams the door in our faces. We stand there, unable to speak.

“So what do we do now?” I ask, my voice unsteady.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

“I don’t know,” Rajan admits.

As we try to figure out our next move, a young man hurries toward us.

“Rajan?” he asks.

“That’s me. Who are you?”

“You don’t remember me? I’m Sonam. I live in this house,” he says, pointing to the same door that was just closed on us. “You saved my life during a climb. Do you remember?”

Rajan’s face softens. “Yes, now I remember.”

For a brief moment, something like relief crosses Sonam’s face. He glances back at the house, then at us, lowering his voice as if the walls can hear.

“You shouldn’t be here,” he murmurs. “My mother is afraid. There are things in this house we do not speak about.”

A chill runs down my spine. I do not wait for him to say more.

I show Mark’s photo to Sonam. He shudders.

“You know him?” Rajan asks.

Sonam hesitates, then nods.

“He is inside our house.”

“What?”

Rajan and I freeze, stunned.

“Why?” I ask. My voice trembles, caught between anger and overwhelming relief at knowing Mark is alive.

“We needed money,” Sonam says quietly. “A man came to us and asked us to hide him. He paid us well. He did not say anything else.”

“Was his name Tenzing?” Rajan asks.

“Yes.”

“That is all we need to know,” Rajan says.

No more words are necessary. We know Mark is alive and who is responsible.

I finally let out the breath I have been holding for far too long.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Epilogue

After several months, I finally hold my brother Mark in my arms again. I have also found Rajan.

Tenzing has been released from the hospital, but he still has to answer for every crime he has committed, including kidnapping.

“There’s just one thing missing,” Mark says to me.

“What?” I ask.

“I want to see Tenzing again.”

“Why?”

“I can’t accept having lost a friend.”

“After everything he did to you?”

“Yes. After everything he did to me. I can’t accept the betrayal of a friend.”

Mark grows quiet after that. His words are not filled with anger, only exhaustion. I realize then that he is not looking for explanations, but for a way to stop carrying the weight of what has happened.

Rajan places a hand on Mark’s shoulder. A simple gesture, steady and silent. It is enough. We all understand where we are going.

The three of us walk toward the prison. It is the same one that once held Rajan. The corridors feel familiar in the worst possible way. Cold walls, controlled silence, the sound of doors closing behind us.

Tenzing is already there, seated on the other side of the glass.

He looks different. Not just thinner, but emptied. When he sees Mark, he straightens, as if instinctively trying to become the person he used to be. It does not work.

“What do you want?” he asks, though his voice betrays him.

Mark steps closer.

“I need to see if you are still the man I knew,” he says.

Tenzing’s hands tighten on the edge of the table.

“I’m not,” he replies after a pause. “And that’s the truth I’ve been avoiding.” He takes a breath. “I tell myself I was doing what I had to do. That everyone would have done the same. But I was lying. I chose the easy way. I chose myself.” His voice falters, but he

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

does not stop. “I don’t expect forgiveness,” he says. “I know I will stay here. I deserve it. But I want you to know that I finally understand what I destroyed.”

Mark listens in silence. His jaw trembles slightly.

“You didn’t just take my freedom,” Mark says. “You took my trust. And I don’t know how to give that back.”

“I know,” Tenzing whispers. “If I could take your place for even one day, I would. But all I can do is stay here and live with what I’ve done.”

The guard announces the end of the visit.

Mark does not move right away. Then, slowly, he leans closer to the glass.

“I forgive you,” he says. “I don’t want to hate you for the rest of my life.”

Tenzing’s eyes fill with tears. He nods, unable to speak.

“That might be the only way I survive this,” he finally says.

Mark places his hand against the glass. After a moment of hesitation, Tenzing does the same, mirroring the gesture.

When we leave the prison, the air outside feels heavy, but breathable.

“I forgive him,” Mark says softly. “Even if the anger is still there. What I feel right now doesn’t matter. What matters is my choice. Only in this way can I find peace.”

Rajan smiles faintly. I take Mark’s hand.

Sometimes forgiveness does not arrive all at once. Sometimes it begins with the decision not to let pain have the last word.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Appendix D. Solutions to Activities

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Chapter 1

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Chapter 2

In this chapter, you learned how to structure an effective PROMPT to best use CHATGPT for Excel, organizing it into three fundamental parts: context, main point, and action. The CONTEXT defines the data and information necessary to frame the problem, the MAIN POINT describes the action required, while the ACTION indicates how to use the result.

ChatGPT can be used as an ASSISTANT to perform various tasks, such as generating formulas for CALCULATION (for example, summing or averaging) or performing search operations with functions like VLOOKUP. This combination of strategies helps you improve productivity, automate complex calculations, and quickly resolve problems in your datasets.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Chapter 3

AUTOFILL IN EXCEL COMPLETES PATTERNS OR DATA AND AI SPEEDS IT.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Chapter 4

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Chapter 5

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Chapter 6

In this chapter, you learned how important proper data PREPARATION is before any analysis.

To support this process, you explored how to structure an effective PROMPT when interacting with COPILOT inside Excel.

Three core operations were covered: identifying and removing DUPLICATES, addressing MISSING VALUES, and applying consistent DATA FORMATTING.

Each of these steps is essential to eliminate noise, correct errors, and ensure that your data can generate reliable INSIGHTS.

For instance, giving Copilot clear and specific instructions allows it to respond more effectively to your needs. Meanwhile, tools such as HIGHLIGHTING RULES and customized formulas support the reliability and consistency of your dataset.

Keep in mind: preparing data isn’t just a technical step; it’s what allows patterns to surface, errors to be minimized, and meaningful conclusions to be drawn with confidence.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Chapter 7

In this chapter, you learned how to use Copilot to explain, generate, and adapt text formulas, handle real-world data errors, and apply conditional logic.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Chapter 8

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Chapter 9

COPILOT GUIDES DATA USE.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Chapter 10

In this chapter, you learned how to use COPILOT in EXCEL to create a complete reporting workflow, from data cleaning to VISUALIZATION, building a SUMMARY, designing a DASHBOARD, generating a LINE CHART, and even automating tasks with a MACRO, all to produce a clear and actionable REPORT.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Chapter 11

AI add-ins enhance Excel by offering specialized tools for data analysis and automation.

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Chapter 12

This chapter shows how Excel, Python, and the OpenAI API work together to build practical AI-driven analysis and classification workflows.

Index

A

accuracy of results, improving with AI in Excel, Why Does Excel Need AI?

action (ChatGPT conversations), The Structure of a Conversation

  setting, Setting the Action

action (main point prompts), Defining the Main Point

activities, solutions to, Chapter 2-Chapter 12

Add all insights to grid button (Copilot), Understanding

  exercise, possible output after clicking, Exercise 3

Add Python cell button, Testing the Python Environment

Add to a new sheet button, Creating Dynamic Summaries

add-ins in Excel, Introducing Add-Ins in Excel-Comparing AI Add-ins and Copilot

  AI-powered add-ins, AI-Powered Add-Ins

  installing, Introducing Add-Ins in Excel

additional details (main point prompt), Defining the Main Point

Agent mode, Introducing and Enabling Copilot in Excel, Removing Duplicates, Explaining Text Formulas

AI (artificial intelligence)

  AI functions in cells, AI-Powered Add-Ins

  architecture of system integrating Excel, Python, and OpenAI, Architecture of Python Add-ins

    Python script calling OpenAI API, OpenAI API

  evaluating AI results, The Benefits of Using AI in Excel

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Excel's need for, Why Does Excel Need AI?-Excel and AI for Data Visualization
reasons to integrate AI tool in Excel, Why Does Excel Need AI?
generating Python code for classification, Generating the Python Code with AI
hallucinations occurring in, Reliability of the Calculated Measures
learning to use, analogy to learning to swim, Exercise 2
music analogy to AI tools, Power GPT
as new imaginary friend, Browsing the Different AI Tools for Excel
output not fully guaranteed, Data Tables
prediction with Python and ChatGPT, Activity 3: Prediction with Python and AI-Activity 3: Prediction with Python and AI
tools for Excel, Browsing the Different AI Tools for Excel-Browsing the Different AI Tools for Excel
troubleshooting classification script with, Using AI for Troubleshooting
using generative AI tools to assist in embedded Python environment, Using Generative AI Tools
AI add-ins for Excel, Landscape of Excel AI Add-Ins-Activity
GPT for Work and GPT for Excel Word, GPT for Work-Exercise 1
Numerous.ai, Numerous.ai-GPT for Work
Power GPT, Power GPT-Power GPT
Sally Suite, Sally Suite-Exercise 2
testing before committing to, Exercise 1
AI add-ins in Excel
comparison with Copilot, Comparing AI Add-ins and Copilot
AI function, Browsing the Different AI Tools for Excel
AI in Excel

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

benefits of using, The Benefits of Using AI in Excel
cases when AI is not useful, When AI Is Not Useful in Excel
roles played by, Summary
AI-aided Formula Editor, Browsing the Different AI Tools for Excel
AI-powered add-ins, AI-Powered Add-Ins
features provided by, AI-Powered Add-Ins
usage-based billing, AI-Powered Add-Ins
Analysis mode, Removing Duplicates, Explaining Text Formulas
Analyst mode, Introducing and Enabling Copilot in Excel
analyst role (ChatGPT), Defining the Main Point
applied fiction, Exercise 2, Before the Nepal Peaks
argument separators in functions, Excel for Data Analysis
Ask Excel Expert
defined, Power GPT
example showing how to use, Power GPT
assistant role (ChatGPT), Defining the Main Point
writing prompt based on, Task 1
Autofill, Autofill-Exercise 3
manually performing in Excel, Autofill-Exercise 3
speeding up using ChatGPT, Exercise 3-Exercise 3
automated data transformations, AI-Powered Add-Ins
AutoSave, Introducing and Enabling Copilot in Excel, Dashboard
average (mean), Measures of Central Tendency
AVERAGE formula, Measures of Central Tendency

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

AVERAGEIF function, Task 3

B

backward fill, Techniques to Deal with Missing Values

Copilot prompt and possible answers for, Techniques to Deal with Missing Values

bar charts, Choose the Right Chart Type

drawing from a PivotTable, Excel for Data Visualization

generating using Excel and ChatGPT, Excel and AI for Data Visualization

provided by Power GPT request, Power GPT

Before the Nepal Peaks, Before the Nepal Peaks-Prologue

activity, creating templates, Activity 2: Creating Templates-Activity 2: Creating Templates

activity, dataset enrichment, Activity 3: Dataset Enrichment-Activity 3: Dataset Enrichment

activity, defining macros, Activity 4: Defining Macros-Activity 4: Defining Macros

activity, getting started, Activity 1: Getting Started-Activity 1: Getting Started

Epilogue, Epilogue

Searching for Truth, Searching for the Truth-Epilogue

The Showdown, The Showdown-Epilogue

biases in AI-generated output, The Benefits of Using AI in Excel

billing, usage based, AI-powered add-ins, AI-Powered Add-Ins

bimodal datasets, Presence of a Dominant Value

blank or missing input values, Describing the Issue

box plots, Visualizing Measures of Dispersion

C

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Caeser cipher activity, Activity

calculation functions

- using ChatGPT to define faster, Calculation

Can I see another insight? button (Copilot), Understanding

categories of data in datasets, Using ChatGPT to Analyze a Dataset

categories, too many, in charts, Simplify and Remove Clutter

- limiting categories, putting in context section of Copilot prompt, Simplify and Remove Clutter

cells

- defined, Basic Excel Concepts
- formatting issues
  - Copilot prompt and answer for dealing with, Techniques to Deal with Data Formatting Issues
  - techniques dealing with, Techniques to Deal with Data Formatting Issues
- misuse of formatting, Data Formatting
- writing Python code directly in, Testing the Python Environment

central tendency (see measures of central tendency)

Chain of Thought (CoT) prompting, Why Does Excel Need AI?

Chart Elements button, Simplify and Remove Clutter

charts, Data Visualization

- Charts group (Insert tab of Ribbon), Excel for Data Visualization
- choosing right chart type, Choose the Right Chart Type
- copies of, in new dashboard, Dashboard
- Copilot choosing best type for you, Choose the Right Chart Type
- Copilot not generating chart itself, Simplify and Remove Clutter

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Copilot writing Python code to produce chart, Using Generative AI Tools

copying from worksheets to dashboard, Dashboard

creating in a PivotTable, error encountered, Excel for Data Visualization

in Excel reports, Data Reporting

foundational principles of effective data visualization, Data Visualization

generated using Python and Copilot, Using Generative AI Tools

Matplotlib for, Embedded Python

chat for general questions (Copilot), Explaining Text Formulas

chat in Copilot, use of, Removing Duplicates

chat panel, integrated, AI-Powered Add-Ins

ChatGPT, Why Does Excel Need AI?-Why Does Excel Need AI?

analyzing dataset with, Using ChatGPT to Analyze a Dataset-Extending to All Questions

ChatGPT for Excel, introduction to, Getting Started with ChatGPT and Excel-Introducing ChatGPT for Excel

conversations with, structure of, The Structure of a Conversation-Setting the Action

creating Excel templates, Creating Templates-Exercise 2

automatically generating template from scratch, Automatically Generating a Template from Scratch-Automatically Generating a Template from Scratch

generating incorrect template, Creating Templates from Scratch

from a photo, Exercise 2

from scratch, Creating Templates from Scratch-Exercise 1

from sketches, Creating Templates from Sketches-Creating Templates from Sketches

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

data enrichment by extracting data from external source, Data Enrichment-Exercise 4

formatting resulting table as Excel table, Data Enrichment

prompt to extract data from the web, Data Enrichment

searching for information on the web, Data Enrichment

writing instructions for extending a table, exercise, Exercise 4

extracting tables based on images, Exercise-Exercise

finding bugs in code, Advanced Operations with ChatGPT

generated output varying by model, Excel and AI for Data Analysis

macro generation in VBA, Using ChatGPT for Macro Generation-Exercise 2

prompting directly in worksheet, Browsing the Different AI Tools for Excel

prompts, using for interactions, Why Does Excel Need AI?

resolving error in VBA macro code generated by, Recording a Macro

resolving VBA compiler error on running macro, Recording a Macro

structuring prompts, exercise, Exercise: Structuring Prompts

using as assistant, writing prompt based on, Task 1

using as housekeeper to handle repetitive tasks, Speeding Up Common Operations with ChatGPT

using as teacher, writing prompt based on, Task 2

using for PivotTables, Using ChatGPT for PivotTables-Using ChatGPT for PivotTables

using model to generalize recorded macro, Recording a Macro

using to speed up Autofill, Exercise 3-Exercise 3

using to speed up creation of formulas and functions in Excel, Task 3-Summary

using with Excel to increase productivity, Why Does Excel Need AI?-Excel and AI for Data Visualization

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

data analysis, Excel and AI for Data Analysis

data management, Excel and AI for Data Management

classification, Browsing the Different AI Tools for Excel, Techniques to Deal with Missing Values, Classification-Exercise 2

generating Python code for with AI, Generating the Python Code with AI

GPT_CLASSIFY function, GPT for Work

using NUM.INFER function, Numerous.ai

Claude, Why Does Excel Need AI?

Clean Data button, Techniques to Deal with Data Formatting Issues

cleanup and normalization (text), Generating Text Formulas

clutter, removing from charts, Simplify and Remove Clutter

code

finding bugs in, Advanced Operations with ChatGPT

column charts, Choose the Right Chart Type

creating using Copilot in Excel, Column Chart-Column Chart

columns, Basic Excel Concepts

renaming using Copilot, Renaming a Column

specifying for Copilot to search for duplicates, Removing Duplicates

splitting using Copilot, Splitting a Column-Exercise 2

comments in VBA, Recording a Macro

communicating with ChatGPT, Getting Started with ChatGPT and Excel

methods for, Introducing ChatGPT for Excel

companion or external tutor, ChatGPT in Excel as, Introducing ChatGPT for Excel

concatenation, Generating Text Formulas

conditional count, Excel for Data Analysis

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

conditional formatting
applied by Copilot to duplicate records, Removing Duplicates
hiding content, Data Formatting
conditional imputation (missing values), Techniques to Deal with Missing Values
Copilot prompt and possible answers for, Techniques to Deal with Missing Values
conditional text logic, Applying Conditional Text Logic
examples of, use cases, and prompts for Copilot, Applying Conditional Text Logic
context
defining for Excel template in prompt, Creating Templates from Scratch
describing in ChatGPT conversations, The Structure of a Conversation
preliminary, providing environment's general settings in Copilot prompt, Adding Slicers
setting in ChatGPT conversation, Setting the Context-Setting the Context
context-main point-action structure, Copilot prompt, Generating Text Formulas
conversations with ChatGPT, Introducing ChatGPT for Excel-Setting the Action
defining the main point, Defining the Main Point-Defining the Main Point
setting the action, Setting the Action
setting the context, Setting the Context-Setting the Context
ConvertToTable subroutine (VBA), Recording a Macro
Copilot
checking for duplicates in dataset, Removing Duplicates
comparison to AI add-ins, Comparing AI Add-ins and Copilot, Summary
data analysis with, Data Analysis with Copilot-Activity
flipping script on approach to tedious tasks, Text Transformation and Manipulation
interpreting statistical results with, Interpreting Statistical Results-Exercise 2

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

missing values, dealing with
practical example, A Practical Example-A Practical Example
suggesting steps to remove duplicates, Removing Duplicates
thinking about how to work with, Adding Slicers
using to understand data, Understanding-Exercise 3
Copilot in Excel, Getting Started with Copilot
benefits and limitations of, Summary, Landscape of Excel AI Add-Ins
blindly trusting, avoiding, Generalizing Data Analysis
calculating central tendency measures, Measures of Central Tendency
replicating stats across worksheets, Replicating Stats Across Worksheets
using single prompt for all measures, Using a Single Prompt for All the Measures
calculating dispersion measures, Measures of Dispersion-Exercise 1
calculating skew and kurtosis, Data Distribution and Shape
calculation of statistical metrics, prompt structure, Introducing Basic Statistics
checking for duplicates in dataset, Removing Duplicates
conditional text logic, applying, Applying Conditional Text Logic
examples, use cases, and prompts for Copilot, Applying Conditional Text Logic
Copilot writing Python code to produce chart
final chart generated, Using Generative AI Tools
creating with, Creating-Exercise 1
crossword puzzle activity, Activity
data analysis with
building advanced PivotTables, Advanced PivotTables-Adding Slicers

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

generalizing data analysis, Generalizing Data Analysis-Generalizing Data Analysis

What-If Analysis, What-If Analysis-Generalizing Data Analysis

data formatting issues, managing

exercises, Exercise 2

new workbook with formatting issues, A Practical Example

possible prompts and answers for, Techniques to Deal with Data Formatting Issues

data reporting with

dashboard, Dashboard-Dashboard

raw data, working with, Raw Data

summary, Defining the Summary

data visualization and reporting with, Data Visualization and Reporting with Copilot

dataset, The Dataset

data visualization with, Data Visualization-Exercise 1

choosing right chart type, Choose the Right Chart Type

PivotTable and line chart generated for objects in space, Data Visualization

prompt structure for, Data Visualization

simplifying and removing clutter, Simplify and Remove Clutter

editing with, Editing-Exercise 2

error handling in text manipulations, Handling Errors-Letting Copilot Detect Problems

Copilot explaining formula for text error handling, Describing the Issue

describing issue to Copilot, Describing the Issue

frequent data error types in Excel and prompts for, Describing the Issue

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

letting Copilot detect problems, Letting Copilot Detect Problems

explaining text formulas, Explaining Text Formulas-Generating Text Formulas

formatting issues, managing

practical example, Techniques to Deal with Data Formatting Issues

using Clean Data button on Ribbon, Techniques to Deal with Data Formatting Issues

generating text formulas, Generating Text Formulas-Exercise 1

common text transformations and prompts for, Generating Text Formulas

exercise, Exercise 1

explaining generated text formula, Generating Text Formulas

prompt structure for, Generating Text Formulas

interpreting statistical results with

presence of a dominant value, Presence of a Dominant Value

reliability of calculated measures, Reliability of the Calculated Measures

introduction to, Introducing and Enabling Copilot in Excel-Introducing and Enabling Copilot in Excel

Copilot panel, Introducing and Enabling Copilot in Excel

options offered by Copilot panel, Introducing and Enabling Copilot in Excel

paid license or Microsoft 365 plan including, Introducing and Enabling Copilot in Excel

data reporting with, Data Reporting

missing values, dealing with

exercise, Exercise 1

possible prompts and Copilot answers for techniques, Techniques to Deal with Missing Values

not allowing automatic generation of external links, Defining the Summary

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

performing creating, editing, and understanding operations, Introducing and Enabling Copilot in Excel

persisting past obstacles with, Exercise 3

using to write Python code, Using Generative AI Tools

generated code, Using Generative AI Tools

mapping between prompt and story structure, Using Generative AI Tools

CopyAllChartsToDashboard subroutine, Dashboard

CoT (see Chain of Thought prompting)

COUNTIF function, Excel for Data Analysis

counting functions

using ChatGPT to define faster, Counting

creating with Copilot, Creating-Exercise 1

column chart, Column Chart-Column Chart

exercise, writing prompt to create PivotTable, Exercise 1

PivotTable, PivotTable-Column Chart

output generated by Copilot, PivotTable

creating, editing, and understanding operations (Copilot), Introducing and Enabling Copilot in Excel

creative desperation, Automatically Generating a Template from Scratch

currency, formatting, Recording a Macro

D

dashboards, Data Reporting, Dashboard-Exercise 2

building in Excel using macros and Copilot, Dashboard

populated with copies of charts and slicer synchronized with all charts, Dashboard

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

prompting Copilot for instructions on copying charts from other worksheets to dashboard, Dashboard

data analysis

about, Data Analysis with Copilot

features provided by Sally Suite, Sally Suite

Pandas for, Embedded Python

using Copilot in Excel, Data Analysis with Copilot-Activity

building advanced PivotTables, Advanced PivotTables-Adding Slicers

generalizing data analysis, Generalizing Data Analysis-Generalizing Data Analysis

What-If Analysis, What-If Analysis-Generalizing Data Analysis

word puzzle activity, Activity

using Excel, Excel for Data Analysis-Excel for Data Visualization

using Excel and AI for, Excel and AI for Data Analysis

Data Analysis feature (Sally Suite), Sally Suite

data cleaning with Numerous.ai, Activity 1: Data Cleaning with Numerous.ai-Activity 1: Data Cleaning with Numerous.ai

data distribution and shape, Introducing Basic Statistics, Data Distribution and Shape-Data Distribution and Shape

calculating using Copilot, Data Distribution and Shape

measures of, use cases, and Excel formulas for, Data Distribution and Shape

data enrichment, Data Enrichment-Exercise 4

extending a table with info from another table, Data Enrichment

extracting data from external source using ChatGPT, Data Enrichment-Exercise 4

data exploration and visualization, Activity 3: Data Exploration and Visualization-Activity 3: Data Exploration and Visualization

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

data extraction using PY function, Activity 2: Data Extraction Using PY()-Activity 2: Data Extraction Using PY()

data formatting (see formatting data)

data management

- using Excel, Excel for Data Management
- using Excel and AI for, Excel and AI for Data Management

data preparation

- about, Data Preparation with Copilot
- formatting data, Data Formatting-Exercise 3
- missing values, dealing with, Dealing with Missing Values-Data Formatting
- removing duplicates, Removing Duplicates-Removing Duplicates
- techniques for, Data Preparation with Copilot
- text completion activity, Activity

data processing, automated, AI-Powered Add-Ins

data reporting (see reporting)

data sharing, disabling for ChatGPT, Setting the Context

data tables, Data Tables

- building two-variable table, exercise, Exercise

data transformations

- automated, with AI-powered add-ins, AI-Powered Add-Ins

data types

- unexpected, in text data, Describing the Issue

data validation, Adding Slicers

data visualization, Activity 3: Data Exploration and Visualization-Activity 3: Data Exploration and Visualization

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

column chart created by ChatGPT from PivotTable, Using ChatGPT for PivotTables

Copilot visualizing measures of dispersion, Visualizing Measures of Dispersion-Exercise 1

creating column chart using Copilot, Column Chart-Column Chart

drawing scatter plot from data using Copilot, Measures of Central Tendency

graph showing sales trends, Exercise 1

using Copilot in Excel, Data Visualization and Reporting with Copilot-Exercise 1

choosing right chart type, Choose the Right Chart Type

Copilot suggesting visual representations of data, Choose the Right Chart Type

dataset, Data Visualization and Reporting with Copilot

prompting Copilot, Data Visualization

simplifying and removing clutter, Simplify and Remove Clutter

using Excel, Excel for Data Visualization

using Excel and AI for, Excel and AI for Data Visualization-Excel and AI for Data Visualization

data, defined, Basic Excel Concepts

DataFrames (Pandas), Testing the Python Environment

Datasaurus dataset, Interpreting Statistical Results

Datasaurus Dozen dataset (Mateika and Fitzmaurice), Measures of Central Tendency

datasets, Basic Excel Concepts

AI difficulties with large datasets, When AI Is Not Useful in Excel

Amazon Top 50 Bestselling Books 2009–2019, extract from, The Structure of a Conversation

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

analyzing using ChatGPT, Using ChatGPT to Analyze a Dataset-Extending to All Questions

dataset to consider, Using ChatGPT to Analyze a Dataset

extending to all questions, Extending to All Questions-Extending to All Questions

extracting category and relevant questions, Extract the Category and Relevant Questions

generating answers, Generate Answers

implementing answers, Implement Answers-Exercise 1

possible questions based on data category, Using ChatGPT to Analyze a Dataset

workflow for, Using ChatGPT to Analyze a Dataset

books dataset for ChatGPT in Excel, Setting the Context

enrichment of, Activity 3: Dataset Enrichment-Activity 3: Dataset Enrichment

hospital patient dataset, Hospital Patient Dataset

enriching with more information using GPT for Excel Word, Enriching a Dataset with More Information

loading into Excel as tables, Excel for Data Management

preparing transactions dataset, Activity 1: Preparing the Dataset-Activity 1: Preparing the Dataset

reading habits dataset, Introducing and Enabling Copilot in Excel

restaurant orders dataset, Advanced PivotTables

sample movie dataset, A Practical Example

space objects, The Dataset

submarine cables in the US, Basic Excel Concepts

fields, Basic Excel Concepts

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

table, data organized in, Basic Excel Concepts

travel dataset, The Dataset

date and time functions

using ChatGPT to define formulas with, Date and Time

dates and times

formatting issues, Data Formatting

Copilot prompt and possible answer for date formats, Techniques to Deal with Data Formatting Issues

fixing date format using Copilot answer, A Practical Example

removing quotation marks from dates using Copilot, A Practical Example

unifying date formats, Techniques to Deal with Data Formatting Issues

decimal separators, Task 3

decluttering charts, Simplify and Remove Clutter

current Copilot version not supporting every type of, Simplify and Remove Clutter

DeepSeek, Sally Suite

default values, Techniques to Deal with Missing Values

Copilot prompt and possible answers for, Techniques to Deal with Missing Values

deletion

Copilot prompt and possible answers for, Techniques to Deal with Missing Values

using for missing values, Techniques to Deal with Missing Values

dependent values, missing or incomplete, Describing the Issue

detecting presence or absence, Applying Conditional Text Logic

dictionary (Scripting.Dictionary), Using ChatGPT for Macro Generation

direct editing tool, ChatGPT in Excel as, Introducing ChatGPT for Excel

dispersion, Introducing Basic Statistics, Using a Single Prompt for All the Measures

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

(see also measures of dispersion)

distribution and shape (see data distribution and shape)

doughnut charts, Choose the Right Chart Type

dropdowns

slicers and, Adding Slicers

using data validation to produce with unique values, Adding Slicers

duplicates, removing from datasets, Removing Duplicates-Removing Duplicates

Copilot checking for duplicates, Removing Duplicates

Copilot suggesting steps to remove duplicates, Removing Duplicates

dialog box to remove duplicates, Removing Duplicates

dynamic prompts, Why Does Excel Need AI?

dynamic summaries, Creating Dynamic Summaries-Adding Slicers

E

editing with Copilot, Editing-Exercise 2

exercise, Exercise 2

renaming columns, Renaming a Column

manually inserting proposed column names, Renaming a Column

translating column names into another language, Renaming a Column

splitting columns, Splitting a Column-Exercise 2

elements of a chart, Simplify and Remove Clutter

equals sign (=), beginning formulas, Basic Functions and Formulas

error handling

in Python code in Excel, Using Generative AI Tools

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

in text manipulations and formulas, Handling Errors-Letting Copilot Detect Problems

approaches to, Handling Errors

describing issue to Copilot, Describing the Issue

exercise, Exercise 2

frequent data error types in Excel and Copilot prompts for, Describing the Issue

letting Copilot detect problems, Letting Copilot Detect Problems

errors generated by AI use, The Benefits of Using AI in Excel

Eurostat, reading habits dataset, Introducing and Enabling Copilot in Excel

Excel

architecture of system integrating Excel, Python, and OpenAI, Architecture of Python Add-ins

basic concepts, Basic Excel Concepts

connecting to Python installation on your computer, Architecture of Python Add-ins

data management with, Excel for Data Management-Excel for Data Management

installing xlwings add-ins, Installing the xlwings Add-ins in Excel

running Python script directly from, Running the Script

slicers in, Adding Slicers

uses of, Why Does Excel Need AI?

Excel AI add-ins (see AI add-ins for Excel)

Excel content, accessing in Python, Testing the Python Environment

Excel Labs, Browsing the Different AI Tools for Excel

explanations, showing in Copilot for formulas, Splitting a Column, Generating Text Formulas

conditional text logic using IF, Applying Conditional Text Logic

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

explaining how to use Goal Seek, Goal Seek

text data error handling formula, Describing the Issue

extraction of data

GPT_EXTRACT function, Browsing the Different AI Tools for Excel

extractions from text, Generating Text Formulas

GPT_EXTRACT function, GPT for Work

F

few-shot prompting, Why Does Excel Need AI?

Field List button, PivotTable tab, Creating Dynamic Summaries

fields

in dataset, submarine cables in the US, Basic Excel Concepts

displaying for PivotTable, Creating Dynamic Summaries

extracting new from existing fields using GPT for Excel Word, Extracting New Fields from Existing Fields

file formats in Excel, Excel for Data Management

files

exporting ChatGPT pivot table as Excel fille, Excel and AI for Data Analysis

file uploads to ChatGPT, Introducing ChatGPT for Excel

loading into ChatGPT, Excel and AI for Data Analysis

filter dropdowns, using in PivotTable fields, Adding Slicers

filtering data before creating charts, Choose the Right Chart Type

final thoughts, Final Thoughts

financial modeling, Why Does Excel Need AI?

Format as Table, Measures of Central Tendency

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

formatting data, Data Formatting-Exercise 3

common problems with in Excel, Data Formatting

defining format for data in Excel template created by ChatGPT, Creating Templates from Scratch

inconsistent formatting, Describing the Issue

techniques to deal with issues, Techniques to Deal with Data Formatting Issues

text transformations, Generating Text Formulas

using Copilot

exercises, Exercise 2

new workbook with formatting issues, Copilot suggestions for, A Practical Example

possible prompts and Copilot answers for techniques, Techniques to Deal with Data Formatting Issues

practical example, Techniques to Deal with Data Formatting Issues

formulas

adding to template worksheet using ChatGPT, Creating Templates from Scratch

drop-down menu in cell containing Python formula, Using Generative AI Tools

using ChatGPT to insert and enhance worksheet, Creating Templates from Scratch

formulas (Excel), Basic Functions and Formulas-Summary

accessing from Ribbon Formulas tab, Excel for Data Analysis

AI-aided Formula Editor, Browsing the Different AI Tools for Excel

Copilot generating, Exercise 1

defining faster using ChatGPT, Basic Functions and Formulas-Exercise

difference between functions and, Basic Functions and Formulas

generated by Copilot to split columns, Splitting a Column-Exercise 2

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Goal Seek and, Goal Seek

Numerous.ai brining AI directly to, Numerous.ai

Power GPT assistance with, Power GPT

forward fill, Techniques to Deal with Missing Values

Copilot prompt and possible answers for, Techniques to Deal with Missing Values

functions

provided by GPT for Work, GPT for Work

provided by Numerous.ai, Numerous.ai

Python, running from Excel buttons or macros, Architecture of Python Add-ins

functions (Excel), Basic Functions and Formulas-Summary

AI functions in cells, AI-Powered Add-Ins

argument separators and function names, Excel for Data Analysis

categories of, Basic Functions and Formulas

defining faster using ChatGPT, Basic Functions and Formulas-Exercise

calculation functions, Calculation

counting functions, Counting

date and time functions, Date and Time

logical functions, Logical

search functions, Search

text manipulation functions, Text Manipulation

difference between formulas and, Basic Functions and Formulas

using ChatGPT for help with, Getting Started with ChatGPT and Excel

G

garbage in, garbage out, When AI Is Not Useful in Excel

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Gemini, Why Does Excel Need AI?, Sally Suite

generalizing data analysis, Generalizing Data Analysis-Generalizing Data Analysis

asking Copilot to propose analyses, Generalizing Data Analysis

proceeding with sales analysis, prompting Copilot, Generalizing Data Analysis

results of running Copilot prompt two more times, Generalizing Data Analysis

Generate capability (Power GPT), Power GPT

advanced mode, Power GPT

generative AI, Breaking the Ice: Introducing AI for Excel

(see also AI)

persisting past obstacles with, Exercise 3

using tools to assist in writing code in embedded Python environment, Using Generative AI Tools

Ghostwriter, Browsing the Different AI Tools for Excel

GitHub repository for this book, Exercise

Amazon Top 50 Bestselling Books 2009–2019 dataset in, The Structure of a Conversation

Books and Info tables in data_enrichment.xlsx example file, Data Enrichment

downloading to local filesystem, Configuring Access to the OpenAI API

Goal Seek, Goal Seek-Data Tables

GPT for Excel Word, GPT for Work-Exercise 1, Summary

adding column to hospital dataset, exercise, Exercise 1

enriching a dataset with more information, Enriching a Dataset with More Information

extracting new fields from existing fields, Extracting New Fields from Existing Fields

main screen and chat feature, GPT for Work

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

GPT for Work, Browsing the Different AI Tools for Excel, GPT for Work-Exercise 1

GPT-4.1 mini, Writing the Script

GPT-4o-mini AI model, Sally Suite

GPT_CLASSIFY function, Browsing the Different AI Tools for Excel, GPT for Work

GPT_EXTRACT function, Browsing the Different AI Tools for Excel, GPT for Work

GPT_TRANLATE function, GPT for Work

gridlines, too many, Simplify and Remove Clutter

H

hallucinations (AI), The Benefits of Using AI in Excel, Data Enrichment

about, Reliability of the Calculated Measures

Copilot interpreting statistical results, Reliability of the Calculated Measures

hidden characters, Data Formatting

clearing, Techniques to Deal with Data Formatting Issues

formatting issues with, Copilot prompt and answers, Techniques to Deal with Data Formatting Issues

highlighting data, Applying Conditional Text Logic

histograms, Visualizing Measures of Dispersion

hospital patient dataset, Hospital Patient Dataset

I

IF function, Handling Errors, Applying Conditional Text Logic

IFERROR function, Handling Errors

inconsistent formatting, Describing the Issue

Insert to new sheet button, Using a Single Prompt for All the Measures

insights into data, showing in Copilot, Understanding

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

insight, defined, Understanding

insights shown by types of charts, Choose the Right Chart Type

text summary of information related to calculated insight, Understanding

instruction-based prompting, Why Does Excel Need AI?

integrated chat panel, AI-Powered Add-Ins

interpolation (for missing values), Techniques to Deal with Missing Values

Copilot prompt and possible answers for, Techniques to Deal with Missing Values

interpretation of statistical results, Interpreting Statistical Results-Exercise 2

determining presence of a dominant value, Presence of a Dominant Value

reliability of calculated measures, Reliability of the Calculated Measures

interquartile range (IQR), Measures of Dispersion

J

junk (in charts), Simplify and Remove Clutter

K

kurtosis, Data Distribution and Shape

different types of, Data Distribution and Shape

L

labels

assigning for specific values in text field, Applying Conditional Text Logic

over-labeling in charts, Simplify and Remove Clutter

LABS.GENERATIVEAI function, Browsing the Different AI Tools for Excel

learning to swim, Exercise 2

leptokurtic distribution, Data Distribution and Shape

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

level of complexity (main point prompts), Defining the Main Point
line charts, Choose the Right Chart Type
links
    Copilot not generating direct links to worksheets, Defining the Summary
    linking worksheets in report workbook, Defining the Summary
logical functions, Logical
low-quality data, difficulty of using AI with, When AI Is Not Useful in Excel

M

machine learning
    models, building and training, Data Preparation with Copilot
Macro-Enabled Workbook (.xlsm format), Recording a Macro
macros, Macros and VBA-Summary, Dashboard
    activating in Excel, Macros and VBA
    Copilot writing macro to add slicer to dashboard, Dashboard
    creating in Excel, quick summary, Dashboard
    defining, Activity 4: Defining Macros-Activity 4: Defining Macros
    enabling, causing disablement of AutoSave, Dashboard
    recording, Recording a Macro
        data formatted as Excel table after, Recording a Macro
        error from VBA compiler when running macro, Recording a Macro
        making macro work correctly for other worksheets, Recording a Macro
        resolving error in code generated by ChatGPT, Recording a Macro
        using ChatGPT model to generalize the macro, Recording a Macro
        workflow to generalize recorded macro, Recording a Macro

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

removing, Dashboard

using ChatGPT to generate macros in VBA, Using ChatGPT for Macro Generation-Exercise 2

VBA code generated by Copilot to copy all charts to dashboard, Dashboard

VBA, generated by Sally Suite, Sally Suite

main function, Writing the Script

main point (ChatGPT conversations), The Structure of a Conversation

defining, Defining the Main Point-Defining the Main Point

contents of main point prompt based on ChatGPT role, Defining the Main Point

roles ChatGPT can play, Defining the Main Point

structure and elements of, Defining the Main Point

defining date and time formulas, Date and Time

defining for calculation functions, Calculation

defining for search functions, Search

Matplotlib, Embedded Python

mean, Measures of Central Tendency

calculating using Copilot, Measures of Central Tendency

comparing with other values in dataset using Copilot, Reliability of the Calculated Measures

measures of central tendency, Measures of Central Tendency-Using a Single Prompt for All the Measures

calculating using Copilot

using single prompt for all measures, Using a Single Prompt for All the Measures

defined, Introducing Basic Statistics

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

mean, median, and mode, use cases and formulas for, Measures of Central Tendency

replicating stats across worksheets using Copilot, Replicating Stats Across Worksheets

using Copilot to calculate, Measures of Central Tendency

measures of dispersion, Introducing Basic Statistics, Measures of Dispersion-Exercise 1

calculating using Copilot, Measures of Dispersion-Visualizing Measures of Dispersion

common measures, use cases, and formulas for, Measures of Dispersion

visualizing using Copilot, Visualizing Measures of Dispersion-Exercise 1

median, Measures of Central Tendency, Reliability of the Calculated Measures

calculating using Copilot, Measures of Central Tendency

mesokurtic distribution, Data Distribution and Shape

metrics

and charts in Excel reports, Data Reporting

(see also charts)

Microsoft 365, Python in Excel, Embedded Python

Microsoft Excel (see Excel)

missing expected text structure, Describing the Issue

missing values, dealing with, Dealing with Missing Values-Data Formatting, Describing the Issue

Copilot replacing missing value with 0, A Practical Example

empty date transformed by Copilot, fixing, A Practical Example

formatting issues, Data Formatting

techniques for, Techniques to Deal with Missing Values

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

using Copilot

exercise, Exercise 1

practical example, A Practical Example-A Practical Example

prompt and possible answers for techniques, Techniques to Deal with Missing Values

mode, Measures of Central Tendency, Presence of a Dominant Value

Star and Bullseye tables not having modes, Replicating Stats Across Worksheets

MODE function, avoiding use of, Replicating Stats Across Worksheets

MODE.MULT function, Measures of Central Tendency

MODE.SINGL function, Measures of Central Tendency, Replicating Stats Across Worksheets

models (AI), choosing for Sally Suite, Sally Suite

multimodal datasets, Presence of a Dominant Value

music, analogy to AI tools, Power GPT

N

natural language instructions, AI-Powered Add-Ins

natural language processing, AI-Powered Add-Ins

negative skew, Data Distribution and Shape

noise in Excel charts, sources of, Simplify and Remove Clutter

normalization (text data), Generating Text Formulas

NUM.AI function, Numerous.ai

extracting only city from addresses, Numerous.ai

NUM.INFER function, Numerous.ai

classifying priorities with, Numerous.ai

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

NUM.WRITE function, Numerous.ai

numbers

formatting issues, Data Formatting

Copilot prompts and answers for techniques addressing, Techniques to Deal with Data Formatting Issues

techniques dealing with, Techniques to Deal with Data Formatting Issues

stored as text, Data Formatting

converting to numbers, Techniques to Deal with Data Formatting Issues

using Copilot to convert to numbers, A Practical Example

Numerous.ai, Browsing the Different AI Tools for Excel, Numerous.ai-GPT for Work, Summary

data cleaning with, Activity 1: Data Cleaning with Numerous.ai-Activity 1: Data Cleaning with Numerous.ai

features provided by, Numerous.ai

functions provided by, Numerous.ai

O

objective (main point prompts), Defining the Main Point

Office Add-ins Marketplace, Introducing Add-Ins in Excel

one-shot prompting, Why Does Excel Need AI?

one-variable data tables, Data Tables

OneDrive

files saved on, using Copilot with, Introducing and Enabling Copilot in Excel

OpenAI

architecture of system integrating Excel, Python, and OpenAI, Architecture of Python Add-ins

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

configuring access to API, Configuring Access to the OpenAI API
Python script calling OpenAI API, OpenAI API
script loading API key, Writing the Script
openpyxl library, Automatically Generating a Template from Scratch
Our World in Data, space objects dataset, The Dataset
outliers, Reliability of the Calculated Measures
overview of the report, Defining the Summary

P

Pandas, Embedded Python
DataFrames, Testing the Python Environment
photos, creating templates from using ChatGPT, Exercise 2
pie charts, Choose the Right Chart Type
pivot tables
building with ChatGPT, Excel and AI for Data Analysis
generated by ChatGPT, formatting as Excel tables, Excel and AI for Data Analysis
PivotTables
advanced, Data Analysis with Copilot-Adding Slicers
adding slicers, Adding Slicers
building in Excel using Copilot, Advanced PivotTables
creating dynamic summaries, Creating Dynamic Summaries-Adding Slicers
dataset, restaurant orders, Advanced PivotTables
building and using, Excel for Data Analysis
comparison to SUMIFS function, Extending to All Questions
creating column chart from using Copilot, Column Chart-Column Chart

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

creating using Copilot, PivotTable-Column Chart
editing generated PivotTable, PivotTable
exercise, writing prompt for, Exercise 1
output generated by Copilot, PivotTable
drawing a bar chart from, Excel for Data Visualization
error encountered when creating chart in, Excel for Data Visualization
generated by Copilot relevant to visualization, Data Visualization
inability of ChatGPT to apply to Excel file independently, Generate Answers
produced using ChatGPT in teacher role, Task 2
provided by Power GPT request, Power GPT
using ChatGPT as teacher to help build, Using ChatGPT for PivotTables-Using ChatGPT for PivotTables
placeholder values, Techniques to Deal with Missing Values
Copilot prompt and possible answers for, Techniques to Deal with Missing Values
formatting issues with, Copilot prompt and answers, Techniques to Deal with Data Formatting Issues
substituting with standardized alternatives or blanks, Techniques to Deal with Data Formatting Issues
platykurtic distribution, Data Distribution and Shape
positive skew, Data Distribution and Shape
Power GPT, Browsing the Different AI Tools for Excel, Power GPT-Power GPT, Summary
core capabilities, Power GPT
example use of Ask Excel Expert, Power GPT
following proposed steps, output produced by, Power GPT
output providing steps to implement request, Power GPT

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

prediction with Python and AI, Activity 3: Prediction with Python and AI-Activity 3: Prediction with Python and AI

predictive modeling

using for missing values, Techniques to Deal with Missing Values

Copilot prompt and possible answers, Techniques to Deal with Missing Values

process automation, Why Does Excel Need AI?

productivity increases using AI in Excel, Why Does Excel Need AI?

example using ChatGPT, Why Does Excel Need AI?

Project Explorer, Dashboard

prompt engineering, Why Does Excel Need AI?

prompts, Why Does Excel Need AI?

communicating with ChatGPT, single prompts and conversations, Introducing ChatGPT for Excel

Copilot creating PivotTable from reading behavior dataset, PivotTable

Copilot detecting issues in text data, Letting Copilot Detect Problems

Copilot error handling in text data, Describing the Issue

Copilot generating text formulas, Generating Text Formulas

Copilot prompt to explain text formula, Explaining Text Formulas

defining, techniques for, Why Does Excel Need AI?

GPT for Excel Word, Extracting New Fields from Existing Fields

Numerous.ai, Numerous.ai

performing Autofill using ChatGPT, Exercise 3

Power GPT, Power GPT

providing to Copilot multiple times, Reliability of the Calculated Measures

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

reformulating question to Copilot in different words, Reliability of the Calculated Measures

structuring, exercise in, Exercise: Structuring Prompts

writing for ChatGPT to generate table, exercise, Exercise 1

writing for use with calculation functions, exercise, Exercise

writing for use with counting functions, exercise, Exercise

writing for use with search functions, exercise, Exercise

writing to generate date and time formulas, exercise, Exercise

writing to generate text formulas, exercise, Exercise

PY function, Embedded Python, Activity 2: Data Extraction Using PY()-Activity 2: Data Extraction Using PY()

testing, Testing the Python Environment

Python, Summary

add-ins, architecture of, Architecture of Python Add-ins-OpenAI API

Excel as frontend, Excel

OpenAI API, OpenAI API

Python, Python

system integrating Excel, Python, and OpenAI, Architecture of Python Add-ins

xlwings, xlwings

advanced analysis and, Techniques to Deal with Missing Values

ChatGPT transforming prompts into, Automatically Generating a Template from Scratch

code provided by Sally Suite, Python Programming from Excel

embedded in Excel, Embedded Python-Exercise 1

dataset, The Dataset

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

enabled Python environment in a cell, Testing the Python Environment

exercise using films dataset, Exercise 1

strict limits of embedded Python environment, Testing the Python Environment

testing the Python environment, Testing the Python Environment

using generative AI tools to assist in writing Python code, Using Generative AI Tools

prediction with Python and AI, Activity 3: Prediction with Python and AI-Activity 3: Prediction with Python and AI

Python, Excel, and AI integration, Python, Excel, and AI-Final Thoughts

cryptogram, solving (activity), Activity

environment setup, Environment Setup-Installing the xlwings Add-ins in Excel

configuring access to OpenAI API, Configuring Access to the OpenAI API

installing xlwings add-ins in Excel, Installing the xlwings Add-ins in Excel

installing Python on your computer, Installing Python on Your Computer

performing classification, Classification-Exercise 2

generating Python code with AI, Generating the Python Code with AI

using AI for troubleshooting, Using AI for Troubleshooting

performing sentiment analysis, Sentiment Analysis-Classification

running the script, Running the Script

writing the script, Writing the Script

Q

QUARTILE.EXC function, Measures of Dispersion

QUARTILE.INC function, caution with, Measures of Dispersion

quotation marks

removing using Copilot instructions, A Practical Example

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

R

ranges, Measures of Dispersion

defined, Basic Excel Concepts

raw data, Data Reporting

adding basic statistics to worksheet, Raw Data

adding legend describing worksheet's columns, Raw Data

Copilot calculating values for each basic statistic, Raw Data

included in worksheet in space_objects.xlsx workbook, Raw Data

redundant series (in charts), Simplify and Remove Clutter

regular expressions produced by Power GPT, Power GPT

Remove Duplicates in Data Tools group, Removing Duplicates

renaming columns using Copilot, Renaming a Column

renaming tables, Excel for Data Management

repetitive tasks in Excel, Speeding Up Common Operations with ChatGPT

reporting, Activity 4: Data Reporting-Activity 4: Data Reporting

using Copilot in Excel, Data Reporting-Summary

completing text using keywords, activity, Activity

dashboard, Dashboard-Dashboard

dataset, Data Visualization and Reporting with Copilot

exercise, Exercise 2

raw data, Raw Data

summary, Defining the Summary

reports, organizing in Excel, Data Reporting

Ribbon in Excel, Excel for Data Management

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

Clean Data button in Data tab, Techniques to Deal with Data Formatting Issues

Formulas tab, Excel for Data Analysis

Formulas tab, Python editor, Testing the Python Environment

Home tab, Copilot icon, Introducing and Enabling Copilot in Excel

Sally Suite tab on, Sally Suite

xlwings tab in, Installing the xlwings Add-ins in Excel

roles ChatGPT can play, Defining the Main Point

rows, Basic Excel Concepts

RunPython command (xlwings), Running the Script

S

Sally Suite, Browsing the Different AI Tools for Excel, Sally Suite-Exercise 2, Summary

output generated from request, Sally Suite

Python code used to compute results, Python Programming from Excel

scatter plots

drawing for Bullseye table, Replicating Stats Across Worksheets

drawing for Dino table, Measures of Central Tendency

drawing for Star worksheet, Replicating Stats Across Worksheets

search functions

using ChatGPT to define faster, Search

Searching for Truth, Searching for the Truth-Epilogue

activity, basic statistics, Activity 2: Basic Statistics-Activity 2: Basic Statistics

activity, data exploration and visualization, Activity 3: Data Exploration and Visualization-Activity 3: Data Exploration and Visualization

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

activity, data reporting, Activity 4: Data Reporting-Activity 4: Data Reporting

activity, preparing the dataset, Activity 1: Preparing the Dataset-Activity 1: Preparing the Dataset

Epilogue, Epilogue

Prologue, Prologue

Select a GPT Function menu, Power GPT

sentiment analysis, Sentiment Analysis-Classification

exercise in, Exercise 2

running the script, Running the Script

directly from Excel, Running the Script

from the terminal, Running the Script

writing the Python script, Writing the Script

shape of data (see data distribution and shape)

Shortcut.AI, Browsing the Different AI Tools for Excel

Show data insights option (Copilot), Understanding

Show explanation option in Copilot for formulas, Splitting a Column

The Showdown, The Showdown-Epilogue

activity, data cleaning with Numerous.ai, Activity 1: Data Cleaning with Numerous.ai-Activity 1: Data Cleaning with Numerous.ai

activity, data extraction using PY function, Activity 2: Data Extraction Using PY()-Activity 2: Data Extraction Using PY()

activity, prediction with Python and AI, Activity 3: Prediction with Python and AI-Activity 3: Prediction with Python and AI

Epilogue, Epilogue

prologue, Prologue-Prologue

single prompts, Introducing ChatGPT for Excel

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

sketches, creating templates from using ChatGPT, Creating Templates from Sketches-Creating Templates from Sketches

skewness, Data Distribution and Shape, Reliability of the Calculated Measures

different types of skews, Data Distribution and Shape

slicers

adding to a PivotTable, Adding Slicers

adding to dashboard, Dashboard

slicer synchronized with all charts in dashboard, Dashboard

special characters, errors from, Describing the Issue

split operations, Generating Text Formulas

splitting columns with Copilot, Splitting a Column-Exercise 2

SQL queries from Power GPT, Power GPT

standard deviation, Measures of Dispersion

statistical imputation (missing values), Techniques to Deal with Missing Values

Copilot prompt and possible answers for, Techniques to Deal with Missing Values

statistics

about, Basic Statistics with Copilot

basic, Activity 2: Basic Statistics-Activity 2: Basic Statistics

basic, introduction to, Introducing Basic Statistics

crossword puzzle activity, Summary

data distribution and shape, Data Distribution and Shape-Data Distribution and Shape

interpreting statistical results, Interpreting Statistical Results-Exercise 2

measures of central tendency, Measures of Central Tendency-Using a Single Prompt for All the Measures

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

measures of dispersion, Measures of Dispersion-Exercise 1

stories, ChatGPT conversions organized as

context, main point, and action, The Structure of a Conversation

practical example, The Structure of a Conversation

styles

applying to Excel template using ChatGPT, Creating Templates from Scratch

subroutines (VBA), Recording a Macro

substring and pattern operations (text), Generating Text Formulas

SUMIFS function, Generate Answers, Extending to All Questions

comparison to PivotTables, Extending to All Questions

summary (in Excel reports), Data Reporting

defining, Defining the Summary

supervised learning of AI, The Benefits of Using AI in Excel

T

Table Design tab, Measures of Central Tendency

tables

ChatGPT formatting data output as Excel tables, Excel and AI for Data Analysis

extending a table with info from another table, Data Enrichment

extracting based on image using ChatGPT, Exercise

formatting raw data as, Excel for Data Management

not formatted by ChatGPT as Excel tables, Automatically Generating a Template from Scratch

tagging by category or keyword, Applying Conditional Text Logic

teacher role (ChatGPT), Defining the Main Point

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

writing prompt based on, Task 2

teacher, Copilot as, Adding Slicers

templates, creating, Creating Templates-Exercise 2, Activity 2: Creating Templates-Activity 2: Creating Templates

using ChatGPT, Creating Templates

auto generation of template from scratch, Automatically Generating a Template from Scratch-Automatically Generating a Template from Scratch

creating templates from scratch, Creating Templates from Scratch-Exercise 1

creating templates from sketches, Creating Templates from Sketches-Exercise 2

text

conditional text logic, applying, Applying Conditional Text Logic

cryptogram, solving (activity), Activity

generating with Copilot, refining prompts step-by-step, Defining the Summary

text data, formatting issues, Data Formatting

Copilot prompts and answers for techniques to address, Techniques to Deal with Data Formatting Issues

managing using Copilot, practical example, A Practical Example

techniques for dealing with, Techniques to Deal with Data Formatting Issues

text formulas

Copilot explaining, Explaining Text Formulas-Generating Text Formulas

Copilot generating, Generating Text Formulas-Exercise 1

categories of text transformation and prompts for, Generating Text Formulas

exercise, Exercise 1

generating using ChatGPT as assistant, Text Manipulation

text manipulation functions

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

completing text with indicated keywords using ChatGPT, activity, Activity

using ChatGPT to define faster, Text Manipulation

title (reports), Defining the Summary

tools (AI), Why Does Excel Need AI?

for use with Excel, Browsing the Different AI Tools for Excel-Browsing the Different AI Tools for Excel

training and improving AI models for ChatGPT, Setting the Context

translations

GPT_TRANSLATE function, GPT for Work

translating column names into another language using Copilot, Renaming a Column

travel dataset, The Dataset

troubleshooting role, ChatGPT, Defining the Main Point

writing prompt based on, Task 3

Trust Center, Macros and VBA

tutor, treating Copilot as, Simplify and Remove Clutter

two-variable data tables, Data Tables

building, exercise, Exercise

U

understanding data using Copilot, Understanding-Exercise 3

elaborating on first insight, buttons for, Understanding

result of clicking Can I see another insight?, Understanding

Show data insights option, Understanding

example of use, Understanding

text summary of information related to calculated insight, Understanding

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

using Add all insights to grid button, exercise, Exercise 3

unexpected data types, Describing the Issue

unimodal datasets, Presence of a Dominant Value

user-defined functions (UDFs), Installing the xlwings Add-ins in Excel

V

validation of data, Adding Slicers

variance, Measures of Dispersion

VBA (Visual Basic for Applications), Macros and VBA-Summary

ChatGPT generating macro in VBA, Using ChatGPT for Macro Generation-Exercise 2

code generated by recording the macro, Recording a Macro

code produced by Power GPT, Power GPT

macros, Dashboard

Sally Suite generating VBA macros, Sally Suite

using ChatGPT model to generalize recorded macro code, Recording a Macro

VBA code generated by Copilot to add slicer to dashboard, Dashboard

VBA code generated by Copilot to copy all charts to dashboard, Dashboard

VBA Editor in Excel, Recording a Macro

examining macro code generated by ChatGPT, Recording a Macro

Verify button (Copilot chat box), Understanding

Visual Basic Editor, Dashboard

Visual Basic for Applications (see VBA)

visualizations, Excel for Data Visualization, Data Visualization

(see also data visualization)

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

VLOOKUP function, Data Enrichment

W

water analogy to editing data, Exercise 2

web server, installing, Installing the xlwings Add-ins in Excel

web version of Excel

adding slicers to PivotTable, prompt to Copilot, Adding Slicers

What-If Analysis, Data Analysis with Copilot, What-If Analysis-Generalizing Data Analysis

Goal Seek, Goal Seek-Data Tables

workbooks

10/space_objects_report.xlsx, Data Reporting

adding insights to using Copilot, Understanding

defined, Basic Excel Concepts

formatting issues in, automatic detection and suggestions for handling from Copilot, A Practical Example

saving as Macro-Enabled Workbook (.xlsm format), Recording a Macro

specifying structure in prompt to ChatGPT, Creating Templates from Scratch

worksheets

in 10/space_objects_report.xlsx workbook, Data Reporting

adding details to in creating template using ChatGPT, Creating Templates from Scratch

ChatGPT adding separate worksheet for each question answered, Extending to All Questions

in Datasaurus Dozen dataset, Measures of Central Tendency

inserting formula into, Measures of Central Tendency

The linked image cannot be displayed. The file may have been moved, renamed, or deleted. Verify that the link points to the correct file and location.

defined, Basic Excel Concepts

linking Summary worksheet to other worksheets in report workbook, Defining the Summary

replicating statistics across, Replicating Stats Across Worksheets

X

xl function, Testing the Python Environment

.xlsm format, Recording a Macro

XLSX file format, Excel for Data Management

XlsxWriter Python library, Automatically Generating a Template from Scratch

xlwings add-in, Architecture of Python Add-ins

essential duties of, xlwings

installing in Excel, Installing the xlwings Add-ins in Excel

RunPython command, Running the Script

Z

zero-shot prompting, Why Does Excel Need AI?

About the Author

Angelica Lo Duca is a researcher at the Institute of Informatics and Telematics of the National Research Council, Italy. She is also an adjunct professor of data journalism at the University of Pisa. Her research interests include data and AI applications to different domains, with a particular focus on data storytelling. She has published over 60 scientific papers at national and international conferences and journals. She often talks at international events, including ACM Tech Talks and the United Nations. She is the author of *Comet for Data Science* (Packt Publishing Ltd, 2022), coauthor of *Learning and Operating Presto* (O'Reilly Media, 2023), *Data Storytelling with Altair and AI* (Manning Publications, 2024), and *Become a Great Data Storyteller* (Wiley, 2025). When she is not working, she loves spending time with her husband, Andrea, and her two children, Giulia and Antonio.

Colophon

The animal on the cover of *Learning Generative AI Tools for Excel* is a Carolina Saddlebags dragonfly (*Tramea carolina*). Carolina Saddlebags is one of the most striking and easily recognizable dragonfly species in North America. It is characterized by its large size and robust build compared to other types of dragonflies, measuring nearly two inches long.

The name *Carolina Saddlebags* comes from its vibrant and distinctive coloring. Male Carolina Saddlebags have deep red coloring across their bodies and wings, with the latter sporting striking black patches. The females, on the other hand, showcase a brownish body with a more amber hue in the wings, which also bear the same black markings.

This dragonfly occupies the southeastern regions of the United States. They prefer warm environments close to bodies of water, such as swamps, lakes, and ponds. They display a notable ability to glide skillfully over the water's surface and are also capable of soaring as high as 23 feet from the ground. Another striking feature is their powerful flight capacity, instrumental in their unique mating rituals: the female often clings onto the male in tandem flight as she submerges her abdomen into the water to deposit her eggs.

The cover illustration is by Monica Kamsvaag based on an antique line engraving from *Histoire Naturelle*. The series design is by Edie Freedman, Ellie Volckhausen, and Karen Montgomery. The cover fonts are Gilroy Semibold and Guardian Sans. The text font is Adobe Minion Pro; the heading font is Adobe Myriad Condensed; and the code font is Dalton Maag's Ubuntu Mono.